JOURNAL OF MATHEMATICS AND PHYSICS

Massachusetts Institute of Technology

Volume III
January–May, 1924

EDITORS
FREDERICK S. WOODS, Professor of Mathematics
HARRY M. GOODWIN, Professor of Physics and Electro-Chemistry
FREDERICK G. KEYES, Professor of Physico-Chemical Research

MANAGING EDITOR
CLARENCE L. E. MOORE, Professor of Mathematics

QA1
J87
V.3

CONTENTS

PAGE

Note on Developable Surfaces in Hyperspace 1
By C. L. E. MOORE

On Ricci's Coefficients of Rotation 7
By JOSEPH LIPKA

Certain Notions in Potential Theory 24
By NORBERT WIENER

Effects of Wood Vanes in Phonograph Horns 52
By W. R. BARSS, J. A. ABBOTT AND H. M. CHATTO

In Memory of Joseph Lipka 63
By NORBERT WIENER

The Coincident Points of Two Algebraic Transformations . . 66
By FRANK LAUREN HITCHCOCK

The Quadratic Variation of a Function and its Fourier Coefficients 72
By NORBERT WIENER

Note on Operational Calculus 95
By V. BUSH

Note on the Quantization of Non-Conditioned-Periodic Systems . 108
By M. S. VALLARTA

A Contribution to the Generalization of a Determinantal Theorem of Frobenius 118
By LEPINE HALL RICE

Continued on Page Three

OHIO STATE
UNIVERSITY

CONTENTS—Continued

PAGE

The Dirichlet Problem 127
By NORBERT WIENER

On Tautochronous Motion 147
By S. D. ZELDIN

The Theory of Testimony 162
By J. S. TAYLOR

Notes on Dynamical Systems Non-Integrable by Separation of Variables and on the Existence of "Unmechanical" Orbits in the Atom 174
By M. SANDOVAL VALLARTA

Note on a Property of Rectilinear Lines of Principal Stress . . 182
By PAUL HEYMANS

Note on the Curl 186
By W. H. INGRAM

Note on a Method of Evaluating the Complex Roots of a Quartic Equation 188
By W. V. LYON

On the Atomic Models of the Alkali Metals 191
By ROBERT BRUCE LINDSAY

On the Dynamic Stresses in Pseudo-Continuous Media 237
By PAUL HEYMANS

315224

NOTE ON DEVELOPABLE SURFACES IN HYPERSPACE

By C. L. E. Moore

1. In ordinary three-space the only type of developable surface is that which is formed by the tangent lines to a twisted curve and a developable surface is transformed into a developable surface by any projective transformation. In hyperspace a surface formed by the tangent lines to a twisted curve is also developable and remains developable under any projective transformation. This is the only kind of ruled developable. There are, however, other types of developables which are not ruled. A few examples of this kind of surface have been given, but otherwise very little is known of them. In this note I wish to discuss in particular those developable surfaces which project on $n-2$ mutually perpendicular three-spaces into ordinary developables. Recently E. B. Wilson and I derived the differential equation of a developable surface[1] in hyperspace.

If the surface is written in the form

$$x=x, \quad y=y, \quad z_i=z_i(x,y), \quad i=1, 2, \ . \ . \ . \ n-2,$$

this equation is

$$\Sigma(r_it_i-s_i^2)[(1+\Sigma p_k^2)(1+\Sigma q_k^2)-(\Sigma p_kq_k)^2-p_i^2(1+\Sigma q_k^2) -q_i^2(1+\Sigma p_k^2)-2p_iq_i\Sigma p_kq_k]$$
$$+\Sigma(r_it_j+r_jt_i-2s_is_j) [(p_iq_j+p_jq_i)\Sigma p_kq_k-p_ip_j(1+\Sigma q_k^2)-q_iq_j(1+\Sigma p_k^2)]=0. \tag{1}$$

where $p_i=\dfrac{\partial z_i}{\partial x}$, $q_i=\dfrac{\partial z_i}{\partial y}$, $r_i=\dfrac{\partial^2 z_i}{\partial x^2}$, $s_i=\dfrac{\partial^2 z_i}{\partial x\partial y}$, $t_i=\dfrac{\partial^2 z_i}{\partial y^2}$,

We pointed out that for a surface to be developable it is not sufficient that $r_it_i-s_i^2=0$, but other relations must be satisfied.

[1] Differential geometry of two dimensional surfaces in hyperspace. Proc. Amer. Acad. Vol. 52, p. 341, 1912.

2. **Four dimensional case.** In case of a surface in four-space equation (1) reduces to

$$(r_1t_1-s_1^2)(1+p_2^2+q_2^2)+(r_2t_2-s_2^2)(1+p_1^2+q_1^2)-(r_1t_2+r_2t_1-2s_1s_1)$$
$$(p_1p_2+q_1q_2)=0, \tag{2}$$

The surfaces we wish to discuss are those for which

$$r_1t_1-s_1^2=0, \quad r_2t_2-s_2^2=0. \tag{3}$$

If these equations are satisfied the surface will project into a developable surface on the three-spaces xyz_1 and xyz_2. Such surfaces can be generated as follows: Take a developable surface in the xyz_1-space and form the three dimensional cylinder v_1 by passing planes through the generators parallel to z_2. Then in the xyz_2-space take any developable and form the cylinder v_2 by passing planes through the generators parallel to z_1. The intersection of v_1 and v_2 will be a surface satisfying (3). Such surfaces will have a double generation of plane curves, for any plane of one cylinder will cut each plane of the other in a point and hence in each plane there will be a curve belonging to the surface. If this surface is developable either

$$r_1t_2+r_2t_1-2s_1s_2=0, \tag{4}$$

or

$$p_1p_2+q_1q_2=0. \tag{5}$$

If (3) and (4) are satisfied, then

$$(r_1t_2+r_2t_1)^2=4s_1^2s_2^2=4r_1t_1r_2t_2,$$

or

$$(r_1t_2-r_2t_1)^2=0.$$

Using (3) and (4) we now have the relations

$$r_1t_2-r_2t_1=0, \tag{6}$$

$$r_1t_2-s_1s_2=0, \tag{7}$$

$$r_1s_2-r_2s_1=0. \tag{8}$$

Equations (6) and (8) are the necessary and sufficient conditions

that the coördinates of the surface satisfy two independent linear partial differential equations of the type

$$A\frac{\partial^2\theta}{\partial x^2}+2B\frac{\partial^2\theta}{\partial x\partial y}+C\frac{\partial^2\theta}{\partial y^2}+D\frac{\partial\theta}{\partial x}+E\frac{\partial\theta}{\partial y}=0. \qquad \text{(A)}$$

When this is the case Segre[2] showed that the surface must either lie in a three-space or be generated by the tangent lines to a twisted curve. Hence *if* (3) *and* (4) *are satisfied the surface is a ruled developable.*

The two independent equations of type (A) which the coördinates of the surface satisfy are

$$r_1\frac{\partial^2\theta}{\partial y^2}-s_1\frac{\partial^2\theta}{\partial x\partial y}=0,$$

$$r_1\frac{\partial^2\theta}{\partial x\partial y}-s_1\frac{\partial^2\theta}{\partial x^2}=0.$$

The other equations which can be formed from (4), (6), (7), (8) are linear combinations of these.

If equations (3) and (5) are satisfied the surface is a nonruled developable. The direction coefficients of the normals to the cylinders V_1 and V_2 at a point of the surface are $(p_1, q_1, 1, 0)$ and $(p_2, q_2, 0, 1)$ respectively. Hence (5) is the necessary and sufficient condition that these normals be perpendicular. If C_1 and C_2 are two plane curves passing through a point P of the surface, the two tangent spaces above are determined by the plane of C_1 and the tangent to C_2 and by the plane of C_2 and the tangent to C_1. Equation (5) is the condition that these spaces be perpendicular.

If (3) is satisfied then p_1 and q_1, p_2 and q_2 are functionally related

$$q_1=f(p_1), \quad q_2=g(p_2).$$

Substituting in (5)

$$p_1p_2+f(p_1)g(p_2)=0 \quad \text{or} \quad \frac{p_1}{f(p_1)}+\frac{g(p_2)}{p_2}=0.$$

[2] Su una classa di superficie degl' iperspasi ecc., Atti di Torino, 1907.

If (4) and consequently (6), (7) and (8) are not satisfied p_1 and p_2 are functionally independent and hence the last equation cannot be satisfied unless

$$\frac{p_1}{f(p_1)} = k = -\frac{g(p_2)}{p_2}, \quad \text{or} \quad \frac{p_1}{q_1} = k, \quad \frac{q_2}{p_2} = -k.$$

Integrating these two equations we have

$$z_1 = F(kx+y), \quad z_2 = G(x-ky).$$

By a simple rotation of the x- and y-axes these become

$$z_1 = F(x), \quad z_2 = G(y).$$

Hence *the only nonruled developables which project into developables on two mutually perpendicular three-spaces are of the form*

$$x = x, \quad y = y, \quad z_1 = F(x), \quad z_2 = G(y),$$

when the axes are properly chosen.

3. **Case of more than four dimensions.** In higher dimensions, as in four dimensions, to have a developable surface it is not sufficient for it to project into a developable surface on the three-spaces xyz_i. From equation (1) it is seen that the surface will be developable if

$$r_i t_i - s_i^2 = 0, \quad r_i t_j + r_j t_i - 2 s_i s_j = 0.$$

All these equations are not independent, for if

$$r_i t_i - s_i^2 = 0, \quad r_1 t_i + r_i t_1 - 2 s_1 s_i = 0 \tag{9}$$

then all the first set must be satisfied, for squaring the second and using the first we obtain

$$r_1 t_i - r_i t_1 = 0, \quad r_1 t_i - s_1 s_i = 0$$

from which we get

$$r_i t_j - r_j t_i = 0, \quad r_i s_j - r_j s_i = 0, \quad r_i t_j - s_i s_j = 0. \tag{10}$$

From (9) and (10) it follows that

$$r_i t_j + r_j t_i - 2 s_i s_j = 0.$$

Hence if (9) are satisfied the surface is developable. From (10)

it is evident that the coördinates of the surface will satisfy the two independent equations

$$r_1\frac{\partial^2\theta}{\partial y^2}-t_1\frac{\partial^2\theta}{\partial x^2}=0,$$

$$r_1\frac{\partial^2\theta}{\partial y^2}-s_1\frac{\partial^2\theta}{\partial x\partial y}=0,$$

and hence from Segre's theorem the surface must be formed by the tangent lines to a twisted curve.

If (9) are satisfied the surface must project on the four-spaces xyz_iz_i into a ruled developable. Thus in a space of five dimensions if a surface projects on the three three-spaces xyz_i into a ruled developable it need not be developable at all, but if it projects on the two four-spaces xyz_iz_i into a ruled developable it must be a ruled developable. If $r_it_i-s_i^2=0$ the surface must have a triple generation of three-space curves or in n-space it will have $n-2$ generations of curves lying in $(n-2)$-space. Using Segre's method it can be shown that if a surface projects on $n-2$ independent four-spaces into a ruled developable it must necessarily be a ruled developable.

From (1) it is seen that the surface will also be developable if

$$r_it_i-s_i^2=0,$$

$$(p_iq_j+p_jq_i)\Sigma p_kq_k-p_ip_j(1+\Sigma q_k^2)-q_iq_j(1+\Sigma p_k^2)=0. \qquad (11)$$

The last equation can be written in the form

$$p_ip_j+q_iq_j+\sum_k(p_iq_k-p_kq_i)(p_kq_j-p_jq_k)=0. \qquad (12)$$

If the projections on the four-spaces xyz_iz_j are non-ruled developables

$$p_ip_j+q_iq_j=0. \qquad (13)$$

From these equations we easily get

$$p_iq_k-p_kq_i=0 \qquad (14)$$

and consequently equations (12) are satisfied. But from (14) we see that z_i and z_k are functionally related so that

$$z_i=f(z_k).$$

Hence the surface will project into a curve on $\frac{(n-2)(n-3)}{2}$ planes and hence must reduce to a curve for $n>5$. We also have on differentiating (13)

$$r_i p_j + r_j p_i + s_i q_j + s_j q_i = 0,$$
$$s_i p_j + s_j p_i + t_i q_j + t_j q_i = 0.$$

Transposing the terms containing s, multiplying and using (11),

$$p_i q_j (r_j t_i - s_i s_j) + p_j q_i (r_i t_j - s_i s_j) = 0,$$

which on using (14) reduces to

$$r_i t_j + r_j t_i - 2 s_i s_j = 0$$

and consequently the surface reduces to a ruled developable. Hence *the only developable surfaces which project on the coördinate spaces xyz_iz_j into the kind of developable discussed in four-space are ruled developables.*

If the surface projects on the four-spaces xyz_iz_i in the type of nonruled developables discussed for four-space then

$$z_j z_k - z_k z_j = 0 \quad j, k \neq 1$$

from which the foregoing shows

$$r_j t_k + r_k t_j - 2 s_j s_k = 0.$$

Consequently from (1) the surface is developable. The same argument used in the case $n=4$ shows the equations of this surface are

$$x=x, \quad y=y, \quad z_1=f(x), \quad z_j=g_j(y) \qquad j=2, 3, \ldots n-2.$$

Hence *if a surface projects on the four-spaces xyz_iz_j into the nonruled type of developable the projections on the four-spaces xyz_kz_j will be ruled developables and the surface will be developable.*

It is easily seen that if the surface projects into a nonruled developable (of the type previously considered) on $n-2$ of the coördinate four-spaces xyz_iz_j so that each of the z-axes is contained in at least one of them, then the surface is developable and the projections on the remaining coördinate spaces xyz_iz_j will be ruled developables. The equations of the surface in this case are

$$x=x, \quad y=y, \quad z_i=f_i(x), \quad z_j=f_j(y),$$
$$i=1, 2, \ldots m, \quad j=m+1, \ldots n-2.$$

ON RICCI'S COEFFICIENTS OF ROTATION

By JOSEPH LIPKA

1. **Introduction.** In a remarkable memoir,[1] "*Dei sistemi di congruenze ortogonali in una varietà qualunque*," Ricci introduced a system of invariants $\gamma_{\alpha\beta h}$ which he called coefficients of rotation. These coefficients are a great aid in the algebra of the absolute calculus, and by their means many geometric theorems are easily derived and the results stated briefly and elegantly.

Ricci shows how these coefficients, which are first defined analytically, may be interpreted kinematically in terms of rotations in the euclidean space in which our fundamental space is imbedded. In a conversation which the author had with Professor Levi-Civita, the latter suggested that these coefficients might be interpreted directly and intrinsically within the fundamental variety by means of the notion of *parallelism* introduced by him.[2] This has been done in Section 2 and Section 3, and the geometric interpretation of $\gamma_{\alpha\beta h}$ is given in the first theorem of the latter section, *viz.:*

Consider a system of n mutually orthogonal congruences of curves in our space, and let $[\alpha]$ and $[\beta]$ be any two distinct congruences of the system and P and P' two neighboring points of a curve of the congruence $[h]$. Consider the angle at P' between the direction determined by the congruence $[\alpha]$ and the parallel to the direction determined at P by the congruence $[\beta]$. Then $\gamma_{\alpha\beta h}$ is the ratio of the cosine of this angle to the arc PP'.

This interpretation allows us immediately to prove several theorems stating the necessary and sufficient conditions for a congruence of curves to be a system of parallel curves or to be a system of geodesics, in terms of the vanishing of certain of the $\gamma_{\alpha\beta h}$. (Section 3).

In Section 2 we disregarded all infinitesimals of higher order than the first. In Section 4 we carry the work to infinitesimals of the second order, and thus derive intrinsically the expression

[1] Memorie della R. Accad. dei Lincei, classe dei Scienze, v. II (1896).

[2] T. Levi-Civita, *Nozione di parallelismo in una varietà qualunque, etc.* Rend. del Circ. Mat. di Palermo, t. XLII. (1917).

for the geodesic curvature and other geometric quantities in terms of the γ's.

In Section 5 and Section 6 by means of the notion of *conformal parallelism*,[3] we introduce a set of invariants $\gamma'_{\alpha\beta k}$ expressed in terms of $\gamma_{\alpha\beta k}$ and an arbitrary point-function ϕ. Analogous theorems to the above are derived for conformally parallel curves, trajectories, and trajectory curvature.

2. **Parallelism and the coefficients of rotation.** As usual, we define a congruence of curves in a space V_n whose fundamental form is[4]

$$ds^2 = \sum_{ik} a_{ik} dx_i dx_k \tag{1}$$

by the equations

$$dx_r = \lambda^r ds \quad (r = 1, 2, \ldots, n) \tag{2}$$

the λ's being functions of the coördinates $x_1, x_2, \ldots, x_n$. Here, the system $\lambda^r (r = 1, 2, \ldots, n)$ form the *contravariant coördinate system* or the *parameters* of the congruence, and their reciprocals

$$\lambda_r = \sum_i a_{ri} \lambda^i \quad (r = 1, 2, \ldots, n).$$

form the *covariant coördinate system* or the *moments*.

We represent by [1], [2], . . . , [n], n mutually orthogonal congruences constituting an orthogonal system in our space, and with $\lambda_1{}^r, \lambda_2{}^r, \ldots, \lambda_n{}^r (r = 1, 2, \ldots, n)$ or with $\lambda_{1/r}, \lambda_{2/r}, \ldots, \lambda_{n/r} (r = 1, 2, \ldots, n)$ the parameters or moments. These congruences determine n mutually orthogonal directions, or an orthogonal ennuple, at each point of V_n. For such ennuples we have the well-known identities

$$\left.\begin{aligned} \sum_r \lambda_\alpha{}^r \lambda_\beta{}^r = \epsilon_{\alpha\beta} = \begin{cases} 0, \text{ if } \alpha \neq \beta \\ 1, \text{ if } \alpha = \beta; \end{cases} \\ \sum_\alpha \lambda_{\alpha/r} \lambda_{\alpha/s} = a_{rs}\,; \sum_\alpha \lambda_\alpha{}^r \lambda_\alpha{}^s = a^{rs}. \end{aligned}\right\} \tag{3}$$

[3] J. Lipka, *On conformal parallelism*, Journ. of Math. and Physics, v. II (1923), pp. 175–194.

[4] Throughout this paper all summations are to extend from 1 to n for the indicated subscripts unless otherwise indicated.

Consider a curve of the congruence $[h]$ and two neighboring points P and P' at a distance ds_h on this curve. Let us designate the orthogonal ennuples formed by the congruences at P and P' by E and E' respectively. If we move the directions in E by *parallelism* to P', these parallels at P' will form a third orthogonal ennuple E'' (since two directions moving by parallelism conserve their concluded angle). Any congruence $[\alpha]$ thus determines three directions: at P the direction α in E, and at P' the directions α' in E' and α'' in E''.

Consider two congruences $[\alpha]$ and $[\beta]$. The direction α has the moments $\lambda_{\alpha/i}$ $(i=1, 2, \ldots, n)$, and the direction α' has the moments

$$\lambda_{\alpha/i}+d\lambda_{\alpha/i}=\lambda_{\alpha/i}+\sum_j \frac{\partial \lambda_{\alpha/i}}{\partial x_j}\,dx_j$$

$$=\lambda_{\alpha/i}+ds_h\sum_j \frac{\partial \lambda_{\alpha/i}}{\partial x_j}\,\lambda_h{}^j. \tag{4}$$

On the other hand, the direction β has the parameters $\lambda_\beta{}^i$ $(i=1, 2, \ldots, n)$ and the direction β'' has the parameters[8]

$$\lambda_\beta{}^i-\sum_{jl}\left\{{j\,l \atop i}\right\}\lambda_\beta{}^l dx_j=\lambda_\beta{}^i-ds_h\sum_{jl}\left\{{j\,l \atop i}\right\}\lambda_\beta{}^l\lambda_h{}^j. \tag{5}$$

It is to be noted that we are carrying our expansions to infinitesimals of the first order only.

Let us designate by $\omega_{\alpha'\beta''}$ the angle between the directions α' and β'', *i.e.*, between the directions given by (4) and (5); then

$$\cos\omega_{\alpha'\beta''}=\sum_i\left[\lambda_{\alpha/i}+ds_h\sum_j\frac{\partial\lambda_{\alpha/i}}{\partial x_j}\lambda_h{}^j\right]\left[\lambda_\beta{}^i-ds_h\sum_{jl}\left\{{j\,l \atop i}\right\}\lambda_\beta{}^l\lambda_h{}^j\right]$$

$$=\sum_i\lambda_{\alpha/i}\lambda_\beta{}^i+ds_h\sum_{ij}\left[\frac{\partial\lambda_{\alpha/i}}{\partial x_j}\lambda_\beta{}^i\lambda_h{}^j-\sum_l\left\{{j\,l \atop i}\right\}\lambda_{\alpha/l}\lambda_\beta{}^l\lambda_h{}^j\right]. \tag{6}$$

Interchanging i and l in the last term, and using (3), (6) may be written

$$\cos\omega_{\alpha'\beta''}=\epsilon_{\alpha\beta}+ds_h\sum_{ij}\left[\frac{\partial\lambda_{\alpha/i}}{\partial x_j}-\sum_l\left\{{j\,i \atop l}\right\}\lambda_{\alpha/l}\right]\lambda_\beta{}^i\lambda_h{}^j. \tag{7}$$

[8] Cf. Reference in footnote 2, p. 7.

The quantity in brackets is the expression for the covariant derivative of $\lambda_{\alpha/i}$, *i.e.* for

$$\lambda_{\alpha/ij}=\frac{\partial \lambda_{\alpha/i}}{\partial x_j}-\sum_l \left\{ {j\ i \atop l} \right\} \lambda_{\alpha/l}; \tag{8}$$

hence (7) takes the form

$$\cos \omega_{\alpha'\beta''}=\epsilon_{\alpha\beta}+ds_h \sum_{ij} \lambda_{\alpha/ij}\lambda_{\beta/}{}^i\lambda_{h/}{}^j. \tag{9}$$

We now define with Ricci[6] the invariants

$$\gamma_{\alpha\beta h}=\sum_{ij} \lambda_{\alpha/ij}\lambda_{\beta/}{}^i\lambda_{h/}{}^j, \tag{10}$$

from which, with the aid of (8), we immediately deduce the identities

$$\gamma_{\alpha\beta h}+\gamma_{\beta\alpha h}=0; \qquad \gamma_{\alpha\alpha h}=0; \tag{11}$$

so that $n^2(n-1)/2$ of the γ's are linearly independent. We finally have

$$\cos \omega_{\alpha'\beta''}=\epsilon_{\alpha\beta}+\gamma_{\alpha\beta h}ds_h. \tag{12}$$

3. **Some geometric consequences.** If the directions α and β are distinct, then, by (3), $\epsilon_{\alpha\beta}=0$, and (12) becomes

$$\gamma_{\alpha\beta h}=\frac{\cos \omega_{\alpha'\beta''}}{ds_h}. \tag{13}$$

(Of course, cos $\omega_{\alpha'\beta''}$ is an infinitesimal since $\omega_{\alpha'\beta''}$ differs from 90° by an infinitesimal.) We have thus arrived at a geometric interpretation of the γ's, which may be stated as follows:

Let $[\alpha]$ *and* $[\beta]$ *be two distinct congruences of our system, and* P *and* P' *be two neighboring points of a curve of a congruence* $[h]$. *Consider the, angle at* P' *between the direction determined by the congruence* $[\alpha]$ *and the parallel to the direction determined at* P *by the congruence* $[\beta]$. *Then,* $\gamma_{\alpha\beta h}$ *is the ratio of the cosine of this angle to the arc* PP'.

As a first consequence of this interpretation, let us find the necessary and sufficient conditions that the curves of the congruence $[\alpha]$ which intersect a curve of the congruence $[h]$ form a system of parallels along this curve. The direction α' must

[6] Cf. Reference in footnote 1, p. 291.

coincide with the direction a''; and since a'' lies in the ennuple E'', it is orthogonal to the other $(n-1)$ directions in E'', hence a' is orthogonal to the same $(n-1)$ directions in E'', and therefore

$$\cos \omega_{a'\beta''}=0, \quad (\beta=1, 2, \ldots, \alpha-1, \alpha+1, \ldots, n),$$

or, by (13),

$$\gamma_{\alpha\beta h}=0, \quad (\beta=1, 2, \ldots, \alpha-1, \alpha+1, \ldots, n). \tag{14}$$

Conversely, if the $(n-1)$ equations (14) are satisfied, then, by (13)

$$\cos \omega_{a'\beta''}=0, \quad (\beta=1, 2, \ldots, \alpha-1, \alpha+1, \ldots, n),$$

and a' is orthogonal to $(n-1)$ directions in E''; but a'' is orthogonal to these same $(n-1)$ directions, hence a' must coincide with a'', and therefore a and a' are parallel directions. Thus —

The necessary and sufficient conditions that the curves of the congruence $[a]$ *which intersect any curve of the congruence* $[h]$ *should form a system of parallels along that curve, are expressed by the* $(n-1)$ *equations* (14).

Corollary: *The necessary and sufficient conditions that the curves of a congruence* $[a]$ *which intersect any curve of any other congruence should form a system of parallels along that curve, are expressed by the* $(n-1)^2$ *equations*

$$\gamma_{\alpha\beta h}=0, \quad (\beta, h=1, 2, \ldots, \alpha-1, \alpha+1, \ldots, n). \tag{15}$$

Corollary: *The necessary and sufficient conditions that the curves of a congruence* $[a]$ *when transported by parallelism along the curves of any other congruence an infinitesimal distance, should remain orthogonal to this latter congruence, are expressed by the* $(n-1)$ *equations*

$$\gamma_{\alpha hh}=0, \quad (h=1, 2, \ldots, \alpha-1, \alpha+1, \ldots, n). \tag{16}$$

Now Ricci has shown[7] that the necessary and sufficient conditions that the congruence $[a]$ should be a *normal congruence*, are expressed by the $(n-1)$ $(n-2)/2$ equations.

$$\gamma_{\alpha\beta h}=\gamma_{\alpha h\beta}, \quad (\beta, h=1, 2, \ldots, \alpha-1, \alpha+1, \ldots, n). \tag{17}$$

[7] Cf. Reference in footnote 1, p. 299.

Combining this with the first corollary above, we may state —

If the curves of a congruence $[a]$ *form a system of parallels along every curve of any other congruence, then* $[a]$ *is a normal congruence.*

The converse is evidently not true.

As another consequence of our interpretation of the γ's, let us find the necessary and sufficient conditions that a congruence $[h]$ should form a congruence of *geodesics*. If in (13) we replace β by h, *i.e.*, if we move the direction h by parallelism from P to P', this equation becomes

$$\gamma_{ahh} = \frac{\cos \omega_{a'h''}}{ds_h} . \tag{18}$$

Now if the curve of the congruence $[h]$ which passes through P and P' is a geodesic, then, by the characteristic property of a geodesic, the parallel at P' to the direction of the curve at P is the tangent direction to the curve at P', *i.e.*, h'' coincides with h'; but h' is orthogonal to the $(n-1)$ other direction in E', hence h'' is likewise, and

$$\cos \omega_{a'h''} = 0, \quad (a = 1, 2, \ldots, h-1, h+1, \ldots, n),$$

or, by (18),

$$\gamma_{ahh} = 0, \quad (a = 1, 2, \ldots, h-1, h+1, \ldots, n). \tag{19}$$

Conversely, if the $(n-1)$ conditions (19) are satisfied, then by (18),

$$\cos \omega_{a'h''} = 0, \quad (a = 1, 2, \ldots, h-1, h+1, \ldots, n),$$

and h'' is orthogonal to $(n-1)$ directions in E'; but h' is orthogonal to these same directions, hence h'' coincides with h', *i.e.* the parallel at P' to the direction of the curve at P is the tangent direction to the curve at P', which is the characteristic property of a geodesic. Hence —

The necessary and sufficient conditions that a congruence $[h]$ *should be a congruence of geodesics, are expressed by the* $(n-1)$ *equations* (19).

4. **The geodesic curvature and other invariants.** In the work of Section 2 and Section 3, we have disregarded infinitesimals of higher order than the first. We shall now find it necessary

to extend the work so as to include infinitesimals of the second order for the case when $\beta=\alpha$, *i.e.*, when we attempt to find the angle at P' between α' and α'', since, by (11), $\gamma_{\alpha\alpha h}$ vanishes identically.

Consider then, as above, two neighboring points P and P' on a curve of the congruence $[h]$. The direction $[\alpha]$ has the moments $\lambda_{\alpha/i}$ and the direction α' has the moments

$$\left.\begin{aligned} &\lambda_{\alpha/i}+A_i ds_h+A'_i ds_h^2+\ldots; \\ &A_i=\frac{d\lambda_{\alpha/i}}{ds_h}=\sum_j \frac{\partial \lambda_{\alpha/i}}{\partial x_j}\lambda_h{}^j; \quad A'_i=\tfrac{1}{2}\frac{dA_i}{ds_h};\ \ldots \end{aligned}\right\} \tag{20}$$

On the other hand, the direction α'' has the moments[8]

$$\left.\begin{aligned} &\lambda_{\alpha/i}+B_i ds_h+B'_i ds_h^2+\ldots; \\ &B_i=\sum_{jl}\left\{{i\,j \atop l}\right\}\lambda_{\alpha/l}\lambda_h{}^j; \quad B'_i=\tfrac{1}{2}\frac{dB_i}{ds_h};\ \ldots \end{aligned}\right\} \tag{21}$$

For the moments given by (20) and (21) we have the identities

$$\left.\begin{aligned} &\sum_{ik}(a^{ik})_{P'}[\lambda_{\alpha/i}+A_i ds_h+A'_i ds_h^2]\,[\lambda_{\alpha/k}+A_k ds_h+A'_k ds_h^2]=1, \\ &\sum_{ik}(a^{ik})_{P'}[\lambda_{\alpha/i}+B_i ds_h+B'_i ds_h^2]\,[\lambda_{\alpha/k}+B_k ds_h+B'_k ds_h^2]=1, \end{aligned}\right\} \tag{22}$$

where $(a^{ik})_{P'}$ is the value of a^{ik} at the point P', *i.e.*,

$$(a^{ik})_{P'}=a^{ik}+\frac{da^{ik}}{ds_h}ds_h+\tfrac{1}{2}\frac{d^2a_{ik}}{ds_h{}^2}ds_h^2+\ldots \tag{23}$$

The infinitesimal angle $d\omega_{\alpha'\alpha''}$ between the directions α' and α'' at P' is given by

$$\cos d\omega_{\alpha'\alpha''}=\sum_{ik}(a^{ik})_{P'}[\lambda_{\alpha/i}+A_i ds_h+A'_i ds_h^2]\,[\lambda_{\alpha/k}+B_k ds_h+B'_k ds_h^2] \tag{24}$$

Using the first identity (22), this becomes

$$\cos d\omega_{\alpha'\alpha''}=1+\sum_{ik}(a^{ik})_{P'}[\lambda_{\alpha/i}+A_i ds_h+A'_i ds_h^2]\,[(B_k-A_k)ds_h+(B'_k-A'_k)ds_h^2]. \tag{24'}$$

8 Cf. Reference in footnote 2, p. 9.

Subtracting the identities in (22), we have

$$2\sum_{ik}(a^{ik})_{P'}[\lambda_{\alpha/i}+A_i ds_k+A'_i ds_k^2]\,[(B_k-A_k)ds_k+(B'_k-A'_k)ds_k^2] + \sum_{ik}(a^{ik})_{P'}[(B_i-A_i)(B_k-A_k)ds_k^2]=0,$$

and introducing this into (24′), the latter becomes

$$\cos d\omega_{\alpha'\alpha''}=1-\tfrac{1}{2}\sum_{ik}(a^{ik})_{P'}(B_i-A_i)(B_k-A_k)ds_k^2;$$

finally, by (23), we get

$$\cos d\omega_{\alpha'\alpha''}=1-\tfrac{1}{2}\sum_{ik}a^{ik}(B_i-A_i)(B_k-A_k)ds_k^2. \tag{25}$$

But by (20) and (21),

$$\sum_{ik}a^{ik}(A_i-B_i)(A_k-B_k)=\sum_{ik}a^{ik}\Big[\sum_j\frac{\partial\lambda_{\alpha/i}}{\partial x_j}\lambda_k{}^j-\sum_{jl}\left\{{i\,j\atop l}\right\}\lambda_{\alpha/l}\lambda_k{}^j\Big]$$

$$\Big[\sum_j\frac{\partial\lambda_{\alpha/k}}{\partial x_j}\lambda_h{}^j-\sum_{jl}\left\{{k\,j\atop l}\right\}\lambda_{\alpha/l}\lambda_h{}^j\Big] \tag{26}$$

and since by (3)

$$a^{ik}=\sum_l\lambda_l{}^i\lambda_l{}^k,$$

(26) takes the form

$$\sum_l\Big[\sum_{ij}\Big(\frac{\partial\lambda_{\alpha/i}}{\partial x_j}-\sum_l\left\{{i\,j\atop l}\right\}\lambda_{\alpha/l}\Big)\lambda_l{}^i\lambda_h{}^j\Big]$$

$$\Big[\sum_{kj}\Big(\frac{\partial\lambda_{\alpha/k}}{\partial x_j}-\sum\left\{{k\,j\atop l}\right\}\lambda_{\alpha/l}\Big)\lambda_l{}^k\lambda_h{}^j\Big]$$

$$=\sum_l\Big(\sum_{ij}\lambda_{\alpha/ij}\lambda_l{}^i\lambda_h{}^j\Big)\Big(\sum_{kj}\lambda_{\alpha/kj}\lambda_l{}^k\lambda_h{}^j\Big)$$

$$=\sum_l\gamma_{\alpha lh}\cdot\gamma_{\alpha lh}=\sum_l\gamma_{\alpha lh}{}^2$$

by (8) and (10). Hence (25) takes the form

$$\cos d\omega_{\alpha'\alpha''}=1-\tfrac{1}{2}ds_h{}^2\sum_l\gamma_{\alpha lh}{}^2,\quad (l\neq\alpha) \tag{27}$$

On the other hand

$$\cos d\omega_{\alpha'\alpha''}=1-\tfrac{1}{2}d\omega_{\alpha'\alpha''}{}^2,$$

and by comparison with (27), we conclude that

$$\left(\frac{d\omega_{a'a''}}{ds_h}\right)^2 = \sum_t \gamma_{ath}^2, (\ t \neq a). \tag{28}$$

We have thus reached a geometric interpretation of the invariant in the right member of (28). This may be stated as follows:

Given any congruence $[a]$ *and two neighboring points* P *and* P' *on a curve of a congruence* $[h]$. *Consider the infinitesimal angle at* P' *between the direction determined by the congruence* $[a]$ *and the parallel to the direction determined at* P *by the same congruence. The invariant* $\sum_t \gamma_{ath}^2$, $(t \neq a)$, *is the square of the ratio of this angle to the arc* PP'.

It is of interest to note that if the above invariant vanishes identically, we must have

$$\gamma_{ath} = 0, \quad (t = 1, 2, \ldots, a-1, a+1, \ldots n),$$

and we have again the necessary and sufficient conditions that the curves of the congruence $[a]$ form a system of parallels along the curves of the congruence $[h]$, as found in Section 3.

As a consequence of the above interpretation, let us consider the infinitesimal angle at P' between the direction of the congruence $[h]$ and the parallel to the direction of the same congruence at P, *i.e.*, between h' and h''. It is merely necessary to set $a = h$ in (28) and obtain

$$\left(\frac{d\omega_{h'h''}}{ds_h}\right)^2 = \sum_t \gamma_{hth}^2 = \sum_t \gamma_{thh}^2, \quad (t \neq h). \tag{30}$$

But the author has shown in an earlier paper[9] that if P and P' are two neighboring points of a curve, the *geodesic curvature* of the curve at P may be intrinsically defined as the ratio of the angle at P' between the direction of the curve and the parallel to the direction of the curve at P, to the arc PP'. Hence the right member of (30) must be the square of the geodesic curvature k of $[h]$ at P. We have thus found intrinsically that

[9] J. Lipka. *Sulla curvatura geodetica delle linee appartenenti ad una varietà qualunque.* Rend. della R. Acc. dei Lincei, v. 31, 1922.

The geodesic curvature κ *of the curve of the congruence* $[h]$ *at the point* P *is given by*

$$\kappa^2 = \sum_t \gamma_{thh}^2, \quad (t \neq h). \tag{31}$$

We again note that if this invariant vanishes identically, we have

$$\gamma_{thh} = 0, \quad (t = 1, 2, \ldots, h-1, h+1, \ldots, n),$$

the necessary and sufficient conditions that the congruence $[h]$ be a geodesic congruence, as found in Section 3.

The quantity κ is the magnitude of a vector whose direction is given by the moments

$$\kappa \mu_{h/r} = \sum_t \gamma_{thh} \lambda_{t/r}, \quad (t \neq h). \tag{32}$$

This direction is evidently orthogonal to the curve h. The curves of the congruence having $\mu_{h/r}$ for moments are said to be the *lines of geodesic curvature* of the congruence $[h]$.

5. **Generalization of $\gamma_{\alpha\beta h}$. Conformal parallelism.** We may generalize the γ's to include in their analytic definition an *arbitrary point function* ϕ $(x_1, x_2, \ldots, x_n)$, by replacing in the development of Section 2, the motion by parallelism from P to P' by that of *conformal parallelism* with characteristic function ϕ.

We have, as before, the orthogonal ennuples E at P and E' at P' formed by the system of congruences. A third orthogonal ennuple E''' at P' now consists of the conformal parallels to the directions of the ennuple E.

Consider two congruences $[\alpha]$ and $[\beta]$. The direction α' has the moments , ,

$$\lambda_{\alpha/i} + ds_h \sum_j \frac{\partial \lambda_{\alpha/i}}{\partial x_j} \lambda_h{}^j,$$

while the direction β''' has the parameters[10]

$$\lambda_\beta{}^i - ds_h \sum_{jl} \left\{ {j\,l \atop i} \right\} \lambda_\beta{}^j \lambda_h{}^l + ds_h \psi^i,$$

[10] Cf. Reference in footnote 3, p. 180.

where

$$\left.\begin{aligned} \psi_i &= \phi^i \sum_h \lambda_h{}^k \lambda_{\beta/k} - \lambda_h{}^i \sum_k \phi^k \lambda_{\beta/k}; \\ \phi_i &= \frac{\partial \phi}{\partial x_i}, \qquad \phi^i = \sum_l a^{il} \phi_l. \end{aligned}\right\} \tag{33}$$

Using the methods and definitions of Section 2, the angle $\eta_{\alpha'\beta'''}$ between α' and β''' is given by

$$\begin{aligned} &\cos \eta_{\alpha'\beta'''} \\ &= \sum_i \left[\lambda_{\alpha/i} + ds_h \sum_j \frac{\partial \lambda_{\alpha/i}}{\partial x_j} \lambda_h{}^j\right]\left[\lambda_\beta{}^i - ds_h \sum_{jl} \left\{{j\,l \atop i}\right\} \lambda_\beta{}^l \lambda_h{}^j + ds_h \psi^i\right] \\ &= \epsilon_{\alpha\beta} + ds_h \sum_{ij} \lambda_{\alpha/ij} \lambda_\beta{}^i \lambda_h{}^j + ds_h \lambda_{\alpha/i} \psi^i. \end{aligned}$$

Introducing the value of ψ^i from (33), we get

$$\begin{aligned} &\cos \eta_{\alpha'\beta'''} \\ &= \epsilon_{\alpha\beta} + ds_h \left[\gamma_{\alpha\beta h} + \sum_i \phi^i \lambda_{\alpha/i} \cdot \sum_k \lambda_h{}^k \lambda_{\beta/k} - \sum_i \lambda_{\alpha/i} \lambda_h{}^i \cdot \sum_k \phi^k \lambda_{\beta/k}\right] \\ &= \epsilon_{\alpha\beta} + ds_h \left[\gamma_{\alpha\beta h} + \epsilon_{h\beta} \sum_i \phi^i \lambda_{\alpha/i} - \epsilon_{h\alpha} \sum_i \phi^i \lambda_{\beta/i}\right]. \end{aligned}$$

Let us now define the invariant

$$\gamma'_{\alpha\beta h} = \gamma_{\alpha\beta h} + \epsilon_{h\beta} \sum_i \phi^i \lambda_{\alpha/i} - \epsilon_{h\alpha} \sum_i \phi^i \lambda_{\beta/i}. \tag{34}$$

Owing to the properties of $\gamma_{\alpha\beta h}$ expressed in (11) and those of $\epsilon_{h\beta}$ and $\epsilon_{h\alpha}$ expressed in (3), we have immediately the identities

$$\gamma'_{\alpha\beta h} + \gamma'_{\beta\alpha h} = 0; \quad \gamma'_{\alpha\alpha h} = 0. \tag{35}$$

We now write

$$\cos \eta_{\alpha'\beta'''} = \epsilon_{\alpha\beta} + \gamma'_{\alpha\beta h} ds_h. \tag{36}$$

6. **Geometric deductions.** If the directions α and β are distinct, then $\epsilon_{\alpha\beta} = 0$, and (36) takes the form

$$\gamma'_{\alpha\beta h} = \frac{\cos \eta_{\alpha'\beta'''}}{ds_h} \tag{37}$$

which gives us a geometric interpretation of the invariants γ'.

Let $[\alpha]$ *and* $[\beta]$ *be two distinct congruences of our system and* P *and* P' *be two neighboring points of a curve of the congruence* $[h]$. *Consider the angle at* P' *between the direction determined by the congruence* $[\alpha]$ *and the conformal parallel to the direction determined at* P *by the congruence* $[\beta]$. *Then* $\gamma'_{\alpha\beta h}$ *is the ratio of the cosine of this angle to the arc* PP'.

If the congruences $[\alpha]$, $[\beta]$, and $[h]$ are all three distinct, then $\epsilon_{h\beta}=0$ and $\epsilon_{h\alpha}=0$, so that (34) becomes

$$\gamma'_{\alpha\beta h}=\gamma_{\alpha\beta h},$$

and hence, $\cos\eta_{\alpha'\beta'''}=\cos\omega_{\alpha'\beta'''}$. We conclude that —

If $[\alpha]$, $[\beta]$, *and* $[h]$ *are any three distinct congruences of our system and* P *and* P' *are two neighboring points of a curve of the congruence* $[h]$, *then the angles at* P' *determined by the direction of the congruence* $[\alpha]$ *with the parallel and the conformal parallel respectively, to the direction of the congruence* $[\beta]$ *at* P, *are equal.*

We may now state, without proof, theorems analogous to those given in Section 3.

1. The necessary and sufficient conditions that the curves of the congruence $[\alpha]$ which intersect any curve of the congruence $[h]$ should form a system of conformal parallels along that curve, are expressed by the $(n-1)$ equations.

$$\gamma'_{\alpha\beta h}=0, \quad (\beta=1, 2, \ldots, \alpha-1, \alpha+1, \ldots, n), \tag{38}$$

or by the equivalent equations

$$\left.\begin{aligned}&\gamma_{\alpha\beta h}=0, \quad (\beta=1, 2, \ldots, h-1, h+1, \ldots, \alpha-1, \alpha+1, \ldots, n);\\ &\gamma_{\alpha h h}+\sum_i \phi^i\lambda_{\alpha/i}=0.\end{aligned}\right\} \tag{38'}$$

2. The necessary and sufficient conditions that the curves of a congruence $[\alpha]$, which intersect any curve of any other congruence should form a system of parallels along that curve, are expressed by the $(n-1)^2$ equations.

$$\gamma'_{\alpha\beta h}=0, \quad (\beta, h=1, 2, \ldots, \alpha-1, \alpha+1, \ldots, n) \tag{39}$$

or by the equivalent equations

$$\left.\begin{aligned}&\gamma_{\alpha\beta h}=0, \quad (\beta, h=1, 2, \ldots, \alpha-1, \alpha+1, \ldots, n), \quad (\beta\neq h),\\ &\gamma_{\alpha h h}+\sum_i \phi^i\lambda_{\alpha/i}=0, \quad (h=1, 2, \ldots, \alpha-1, \alpha+1, \ldots n).\end{aligned}\right\} \tag{39'}$$

3. The necessary and sufficient conditions that the curves of the congruence $[\alpha]$ when transported by conformal parallelism along the curves of any other congruence an infinitesimal distance, should remain orthogonal to this latter congruence, are expressed by the $(n-1)$ equations.

$$\gamma'_{\alpha hh}=\gamma_{\alpha hh}+\sum\phi^i\lambda_{\alpha/i}=0,\ (h=1, 2, \ldots, \alpha-1, \alpha+1, \ldots, n). \quad (40)$$

4. The necessary and sufficient conditions that the congruence $[\alpha]$ should be a normal congruence, are expressed by the $(n-1)$ equations.

$$\gamma'_{\alpha\beta h}=\gamma'_{\alpha h\beta},\quad (\beta, h=1, 2, \ldots, \alpha-1, \alpha+1, \ldots, n) \quad (41)$$

This result follows immediately by adding the term $\epsilon_{h\beta}\sum_i\phi^i\lambda_{\alpha/i}$ to each member of equation (17).

5. If the curves of a congruence $[\alpha]$ form a system of conformal parallels along the curves of any other congruence which they intersect, then $[\alpha]$ is a normal congruence.

Let the curves of a congruence be *trajectories*, *i.e.*, curves which render

$$\int e^\phi ds = \text{minimum}.$$

Such a curve has the characteristic property that its directions at all points are conformally parallel.[11] It follows that —

6. *The necessary and sufficient conditions that a congruence* $[h]$ *be a congruence of trajectories, are expressed by the* $(n-1)$ *equations*

$$\gamma'_{ahh}=\gamma_{ahh}+\sum\phi^i\lambda_{a/i}=0,\ (a=1, 2, \ldots, h-1, h+1, \ldots, n). \quad (42)$$

7. **The relative curvature and other invariants.** In Section 5 we have disregarded infinitesimals of higher order than the first. If we wish to find the angle at P' between α' and α''' *i.e.*, to consider the case when $\beta=\alpha$, and therefore, by (35) $\gamma'_{\alpha\alpha h}=0$ identically, we must extend our work to include infinitesimals of the second order, as we have done in Section 4.

Consider, then, two neighboring points P and P' on a curve

[11] Cf. Reference in footnote 3, p. 182.

of the congruence $[h]$. The direction α has the moments $\lambda_{\alpha/i}$ and the direction α' has the moments

$$\left.\begin{aligned} &\lambda_{\alpha/i}+A_i ds_h+A'_i ds_h^2+\ \dots\ ; \\ &A_i=\frac{d\lambda_{\alpha/i}}{ds_h}=\sum_j \frac{\partial \lambda_{\alpha/i}}{\partial x_j}\lambda_h{}^j;\ \ A'_i=\tfrac{1}{2}\frac{dA_i}{ds_h};\ \dots \end{aligned}\right\} \tag{43}$$

On the other hand, the direction α''' has the moments[13]

$$\left.\begin{aligned} &\lambda_{\alpha/i}+B_i ds_h+B'_i ds_h^2+\ \dots\ ; \\ &B_i=\sum_{jl}\left\{{i\,j \atop l}\right\}\lambda_{\alpha/l}\lambda_h{}^j+\psi_i;\ \ B'_i=\tfrac{1}{2}\frac{dB_i}{ds_h};\ \dots \end{aligned}\right\} \tag{44}$$

Proceeding as in Section 4, we find for the infinitesimal angle $d\eta_{\alpha'\alpha'''}$ between α' and α''',

$$\cos d\eta_{\alpha'\alpha'''}=1-\tfrac{1}{2}\sum_{ik}a^{ik}(B_i-A_i)(B_k-A_k)ds_h^2.$$

By (43) and (44), this becomes

$$\cos d\eta_{\alpha'\alpha'''}=1-\tfrac{1}{2}ds_h^2\sum_t(\gamma_{\alpha th}-\sum_i\psi_i\lambda_t{}^i)(\gamma_{\alpha th}-\sum_k\psi_k\lambda_{t/k}),\ (t\neq\alpha).$$

But, by (33),

$$\begin{aligned}&\sum_i\psi_i\lambda_t{}^i \\ &=\sum_i\psi^i\lambda_{t/i}=\sum_k\lambda_h{}^k\lambda_{\alpha/k}\cdot\sum_i\phi^i\lambda_{t/i}-\sum_i\lambda_h{}^i\lambda_{t/i}\cdot\sum_k\phi^k\lambda_{\alpha/k} \\ &=\epsilon_{h\alpha}\sum_i\phi^i\lambda_{t/i}-\epsilon_{ht}\sum_i\phi^i\lambda_{\alpha/i},\end{aligned}$$

and, hence,

$$\gamma_{\alpha th}-\sum_i\psi_i\lambda_t{}^i=\gamma'_{\alpha th}+\epsilon_{ht}\sum_i\phi^i\lambda_{\alpha/i}-\epsilon_{h\alpha}\sum_i\phi^i\lambda_{t/i}=\gamma'_{\alpha th},$$

by (34). Thus, finally,

$$\cos d\eta_{\alpha'\alpha'''}=1-\tfrac{1}{2}ds_h^2\sum_t\gamma'^2_{\alpha th},\quad (t\neq\alpha). \tag{45}$$

But

$$\cos d\eta_{\alpha'\alpha'''}=1-\tfrac{1}{2}(d\eta_{\alpha'\alpha'''})^2,$$

[13] Cf. Reference in footnote 3, p. 180.

and so we may write

$$\left(\frac{d\eta_{a'a'''}}{ds_h}\right)^2 = \sum_t \gamma'^2_{ath}, \quad (t \neq a), \tag{46}$$

giving us a geometric interpretation of the invariant in the right member. Thus —

Given any congruence $[a]$ *and two neighboring points* P *and* P' *on a curve of a congruence* $[h]$. *Consider the infinitesimal angle at* P' *between the direction determined by the congruence* $[a]$ *and the conformal parallel to the direction determined at* P *by the same congruence. The invariant* $\sum_t \gamma'^2_{ath}$, $(t \neq a)$, *is the square of the ratio of this angle to the arc* PP'.

We note that if the above invariant vanishes identically, we have again the necessary and sufficient conditions that the curves of the congruence $[a]$ form a system of conformal parallels along the curves of the congruence $[h]$, as expressed in Section 6.

As an application of the above interpretation, we may consider the infinitesimal angle at P' between the direction of the congruence $[h]$ and the conformal parallel to the direction of the same congruence at P, *i.e.*, between h' and h'''. It is only necessary to set $a = h$ in (46) and obtain

$$\frac{d\eta_{h'h'''}}{ds_h} = \sum_t \gamma'^2_{hth} = \sum_t \gamma'^2_{thh}, \quad (t \neq h). \tag{47}$$

Now, if we draw the trajectory through P in the direction PP', h''', the conformal parallel to h, will give the direction of the tangent line to this trajectory at P'. Thus the ratio $d\eta_{h'h'''}/ds_h$ is the ratio of the angle at P' between the h curve and the trajectory, to the arc PP'. In a previous paper,[13] the author designated the ratio (or the limit of the ratio) of the angle at P' between any two curves c_1 and c_2 having an element PP' in common, to the arc PP', by the term *relative curvature* of c_1 with respect to c_2. If the curve c_2 is a *geodesic*, this gives us the ordinary *geodesic curvature* of the curve c_1; and if the curve c_2 is a *trajectory*, we may similarly designate the above ratio by term *trajectory curvature* of the curve c_1.

[13] J. Lipka, *On the relative curvature of two curves in Vn*, Bull. Am. Math. Soc., Vol. 29, October, 1923.

Hence, we conclude —

The trajectory curvature κ' of the curve of the congruence $[h]$ at the point P is given by

$$\kappa'^2 = \sum_t \gamma'^2_{thh}, \quad (t \neq h). \tag{48}$$

The quantity κ' is the magnitude of a vector whose directions are given by the moments

$$\kappa' \mu'_{h/r} = \sum_t \gamma'_{thh} \lambda_{t/r}, \quad (t \neq h), \tag{49}$$

This direction is evidently orthogonal to the curve h. The curves of the congruence having $\mu_{h/r}$ for moments are said to be the *lines of trajectory curvature* of the congruence $[h]$.

If κ is the geodesic curvature of the curve h, then, by (32) the direction of the geodesic curvature vector of h is given by

$$\kappa \mu_{h/r} = \sum_t \gamma_{thh} \lambda_{t/r}, \quad (t \neq h).$$

If κ'' is the geodesic curvature of the trajectory T tangent to the curve h at P, then, by (32), the direction of the geodesic curvature vector of T is given by

$$\kappa'' \mu''_{T/r} = \sum_t \gamma_{tTT} \lambda_{t/r} = -\sum_t \left(\sum_i \phi^i \lambda_{t/i} \right) \lambda_{t/r}, \quad (t \neq T),$$

where we have used the characteristic equation (42) for a trajectory system.

Finally, if κ' is the trajectory curvature of the curve h, then, by (49), the direction of the trajectory curvature vector of h is

$$\kappa' \mu'_{h/r} = \sum_t \gamma'_{thh} \lambda_{t/r}, \quad (t \neq h).$$

These three vectors are perpendicular to h at P. But, by (34),

$$\gamma'_{thh} = \gamma_{thh} + \sum_i \phi^i \lambda_{t/i}, \quad (t \neq h);$$

hence,

$$\kappa' \mu'_{h/r} = \kappa \mu_{h/r} - \kappa'' \mu''_{T/r}, \tag{50}$$

and we may state —

The three directions at a point P — the geodesic curvature vector

of a curve h, the geodesic curvature vector of its tangent trajectory and the trajectory curvature vector of h — lie in a same flat pencil of directions.

If we multiply (50) by the corresponding relation

$$\kappa'\mu'_{h/}{}^{r} = \kappa\mu_{h/}{}^{r} - \kappa''\mu''_{T/}{}^{r},$$

and sum on r, we find, since

$$\sum_{r}\mu_{h/r}\mu_{h/}{}^{r} = 1, \quad \sum_{r}\mu_{h/r}\mu_{T/}{}^{r} = \cos\theta,$$

the relation

$$\kappa'^{2} = \kappa^{2} + \kappa''^{2} - 2\kappa\kappa''\cos\theta. \tag{51}$$

Combining this with (46), we may state —

The invariant $\sum_{t}\gamma'^{2}_{thh}$, $(t \neq h)$, *represents the sum of the squares of the geodesic curvatures at a point P of the curve h and its tangent trajectory minus twice the product of these curvatures and the cosine of the angle between their geodesic curvature vectors.*

November 21, 1923.

CERTAIN NOTIONS IN POTENTIAL THEORY[1]

By NORBERT WIENER

Introduction.

§1. The relative weights of two point-sets.

§2. The generalized Dirichlet problem.

§3. The distribution of charge on an isolated conductor.

§4. Capacity, and a sufficient condition for the solubility of the Dirichlet problem.

5. The *im Kleinen* character of the Dirichlet problem.

Introduction. The Dirichlet problem — of determining a harmonic function assuming a given continuous set of values continuously on the boundary of a region R — does not always admit of a solution for every region R. For example, if we use polar coördinates and take as our R the region $0 < r < 1$, we shall find that there is no harmonic function assuming the boundary values 0 for $r=1$ and 1 for $r=0$. In order to see why this is so, let us replace R by the slightly modified region $\epsilon < r < 1$, and our boundary conditions by the stipulation that the harmonic function shall vanish for $r=1$, and shall equal 1 for $r=\epsilon$. The Dirichlet problem then becomes soluble, and has as its answer $u(r, \theta) = \dfrac{\log r}{\log \epsilon}$. Now, we have for every point in R

$$\lim_{\epsilon \to 0} u(r, \theta) = 0.$$

It thus appears that by modifying the boundary slightly and then letting the harmonic function determined on the interior of the new boundary approach its limit as the modification of the boundary becomes progressively less and less, we get a definite

[1] The author wishes to express his thanks to Dr. J. L. Walsh and to Prof. O. D. Kellogg, both of Harvard University, with whom he has discussed matters in this paper, and to whom he owes much in the way of suggestion and stimulus. In spirit, in particular in the importance given to notions such as charge and capacity and in the use of Stieltjes integrals, this paper has much in common with one of Professor Kellogg. (*An Example in Potential Theory*, Proc. Amer. Acad. Arts and Sciences, Vol. 58, No. 14, June, 1923).

harmonic function as a limit — namely 0 — which fails, however, to assume continuously the assigned boundary values at $r=0$. The origin is, as it were, a part of the boundary of zero weight, and a boundary condition at the origin is null and void.

It will be seen, then, that the Dirichlet problem fails to be soluble in this particular case, not by the non-existence of a harmonic function corresponding to the boundary conditions, but by the fact that this harmonic function fails to assume continuously the desired value at the origin. We shall see that whenever the Dirichlet problem fails to be soluble, it is possible to determine a perfectly general function corresponding to our boundary conditions. In more precise language, our theorem reads:

THEOREM I. *Let R be a connected, bounded set of points in n-space, containing no point of its boundary. Let C be its boundary. Let $U(P)$ be any function continuous on the set of points C. Then*

1. *There is a function $v(P)$ continuous over $R+C$ and reducing to $U(P)$ on C;*

2. *There is a sequence $\{\Gamma_n\}$ of boundaries over the interior of which the Dirichlet problem is soluble for any continuous boundary condition, each contained in the interior of its successor, and together containing in their interiors every point of R;*

3. *The function $u_n(P)$ corresponding to the boundary conditions $v(P)$ over Γ_n and harmonic within Γ_n tends uniformly to a harmonic limit over any closed set of points interior to R as n increases without limit;*

4. *This limit, which we shall term $u(P)$, is independent of the choice of $v(P)$ and $\{\Gamma_n\}$.*

A precisely similar theorem will hold for the exterior Dirichlet problem. A particularly interesting case of the exterior Dirichlet problem arises when the assigned values on the boundary are constant over the entire boundary, while the assigned behavior of the desired harmonic function at infinity is the same as that of the potential due to a single charge. Physically, this problem is that of determining the potential due to a distributed charge on a conductor. However a charge on a conductor will not always distribute itself in such a way as to possess a surface-density.

Hence it will not always be possible to find a surface distribution of charge in the ordinary sense which will lead to a solution of this exterior Dirichlet problem. If we replace the ordinary notion of surface distribution, leading to a potential represented by a surface integral in the three-dimensional case, by a surface distribution corresponding to a three-dimensional Stieltjes integral, this difficulty will be removed, and we shall obtain a particular case of:

THEOREM II. *Let R be any bounded set of points in n-space $(n>2)$ inside the rectangular parallelepiped with opposite vertices $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$. Let $u(P)$ be a harmonic function with no singularities except on R, vanishing at infinity, and corresponding to boundary values 1 over R in the sense of Theorem I. Then over every point Q exterior to R we shall have*

$$u(Q)=\int_{a_1}^{b_1}\int_{a_2}^{b_2}\cdots\int_{a_n}^{b_n}(\overline{PQ})^{2-n}\,dM(P)$$

where $M(P)$ is an increasing function of the cartesian coördinates of P and the integral is in the sense of Daniell[3] *an n-dimensional Stieltjes integral. Moreover, if the parallelepiped with opposite vertices $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ is entirely exterior to R, then*

$$\int_{x_1}^{y_1}\cdots\int_{x_n}^{y_n}dM(P)=0.$$

An analogous theorem will be proved in the two-dimensional case.

Theorem II makes it possible to define the electrostatic capacity of a set of points. This is the total charge corresponding to a unit potential over the set. In other words, it is

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}dM(P).$$

The capacity of a portion of the boundary of a region gives a more precise definition than any yet given of the amount of boundary in that portion in the sense which is relevant with regard to the solubility of the Dirichlet problem. We shall establish:

[3] P. J. Daniell, Annals of Mathematics, Vol. 21 (1921), p. 30.

THEOREM III. *Let R, C, U, and u be defined as in Theorem I. Let Q be a point of C. Draw an n-sphere S with variable radius r and Q for center. Let the capacity of that part of C lying in or on the surface of S be C_1. Let the capacity of S be C_2. Let the outer Lebesgue measure of that part of S within R, in terms of S itself, be L. Then if there are a denumerable number of values of r tending to zero for which either*

$$L < N < 1,$$

or

$$C_1 > MC_2 > 0,$$

M and N being independent of R, then

$$\lim_{P \to Q} u(P) = U(Q).$$

In particular, if this condition is fulfilled for every point Q of C then the Dirichlet problem is soluble for any continuous boundary condition on C.[3]

As far as the knowledge of the author goes, this condition is satisfied by every example of an explicit solution of the Dirichlet problem yet described in the literature. It is not the opinion of the author, however, that it is a necessary condition for the solubility of the Dirichlet problem.

A further matter which comes within the scope of this paper is the effect of boundary conditions at a remote point in modifying the continuity of the $u(P)$ of Theorem I at a given point of the boundary. We shall prove:

THEOREM IV.[4] *Let R be a connected, bounded set of points in n-space, containing no point of its boundary. Let Q be a point of its boundary. Let S be a sphere with Q as center. Let R' be another connected, bounded set of points and let the portions of R and of R' within S concide. Let C' be the boundary of R'.*

[3] A theorem of content implying a part of the results of Theorem III is contained in Phillips and Wiener, *Nets and the Dirichlet Problem*, Jour. Math. Phys. M. I. T., Vol 2, pp. 105–124. Cf. also G. E. Raynor, *Dirichlet's Problem*, Annals of Mathematics, Second Series, Vol. 23, pp. 183–198.

[4] A part of the content of this theorem has been enunciated by Prof. O. D. Kellogg, in a forthcoming paper.

Let $u(P)$, defined as in Theorem I, satisfy the condition

$$\lim_{P \to Q} u(P) = U(Q)$$

for any function U continuous over C. Let $V(P)$ be any function continuous over C'. Let $v(P)$ be the function harmonic over R' corresponding to $V(P)$ after the fashion indicated in Theorem I. Then

$$\lim_{P \to Q} v(P) = V(Q).$$

§1. **The relative weights of two point-sets.** Let H and K be two closed bounded sets of points in the plane. Let the lower bound of the distance between a point of H and a point of K be the positive number d. Let $\{ H_n \}$ be a set of boundaries, each of which divides the plane into an inside and an outside region, each containing all its successors, with only the points of H common to all. Let $\{ K_n \}$ be a similar sequence about K. Let the exterior Dirichlet problem be soluble for any H_n or K_n, in the sense that it is possible to find a harmonic function over the exterior of the portion of boundary in question, assuming arbitrary given continuous values continuously on the boundary, and bounded at infinity. It will always be possible to construct sequences $\{ H_n \}$ and $\{ K_n \}$ for any H and K. For example, we may first divide the plane into little squares with corners of the form $(i/2^n, j/2^n)$, and take H_n as the boundary of the set of points lying within such of the squares as contain points of H.

It may be shown by a ready application of the alternating process that if m and n are chosen so large that H_m and K_n do not intersect, the exterior Dirichlet problem is soluble for the region excluded by H_m and K_n. Let the harmonic function assuming the values 1 on H_m and $f(P)$ on K_n be $u_{mn}(f; P)$. Then as m increases, the sequence of functions $u_{mn}(f; P)$ is non-increasing. Hence by a theorem of Harnack[5] the sequence $u_{mn}(f; P)$ converges uniformly to a harmonic limit, which we shall call $u_n(f;P)$, over any closed set of points outside H and K. Similarly, let the harmonic function assuming the values $g(P)$ on H_m and 0 on K_n be $\overline{u}_{mn}(g;P)$. Then as n increases, the sequence of func-

[5] Cf. Osgood, *Funktionentheorie*, p. 615.

tions $\bar{u}_{mn}(g;P)$ is non-decreasing. By the same theorem of Harnack, the sequence $\bar{u}_{mn}(g;P)$ converges uniformly to a harmonic limit. which we shall call $\bar{u}_m(g;P)$, over any closed set of points outside H_m and K.

Without any essential restriction, we may suppose that H_1 and K_1 are each exterior to the other. It follows from the considerations which we have just presented that from some value of m on, the functions $u_{mn}(0;P)$ will differ from $u_n(0;P)$ on H by less than ϵ, while the functions $\bar{u}_{mn}(1;P)$ will from some value of n on differ from $\bar{u}_m(1;P)$ on K by less than ϵ. Moreover, there will be a positive number b less than 1, such that for all m and n.

$$\left.\begin{aligned} u_{m1}(0;P) &= \bar{u}_{m1}(1;P) > 1-b \text{ on } H_1; \\ u_{1n}(0;P) &= \bar{u}_{1n}(1;P) < b \quad \text{on } K_1. \end{aligned}\right\} \tag{1}$$

Now let us obtain a function harmonic outside H_m and K_n, assuming the value 1 on H_m and 0 on K_n, by the alternating process. Let us write $v_0(P)$ for $u_{m1}(0;P)$ and $w_0(P)$ for $u_{1n}(0;P)$. Let $w_k(P)$ be the harmonic function which assumes the value 0 on K_n and the same value as $v_{k-1}(P)$ on H_1. Let $v_k(P)$ be the harmonic function which assumes the value 1 on H_m and the same value as $w_{k-1}(P)$ on K_1. We shall have:

$$\left.\begin{aligned} |\, v_1(P)-v_0(P) \;| &< b, \\ |\, w_1(P)-w_0(P) \;| &< b, \\ |\, v_2(P)-v_1(P) \;| &< b^2, \\ |\, w_2(P)-w_1(P) \;| &< b^2, \\ &\cdots\cdots \\ |\, v_k(P)-v_{k-1}(P) \;| &< b^k, \\ |\, w_k(P)-w_{k-1}(P) \;| &< b^k, \\ &\cdots\cdots \end{aligned}\right\} \tag{2}$$

Hence the series $\sum_1^\infty [v_k(P)-v_{k-1}(P)]$ and $\sum_1^\infty [w_k(P)-w_{k-1}(P)]$ are uniformly and absolutely convergent to a harmonic limit for every P for which they are defined and for every sufficiently late m and n. Over the region excluded by both H_1 and K_1 the functions $v_k(P)$ and $w_k(P)$ approach the same boundary values, and hence the same values everywhere. The limit approached by $v_k(P)$ and $w_k(P)$ as k becomes infinite we shall designate over

the entire region exterior to H_1 and K_1 as $F_{mn}(P)$. That is,

$$\left.\begin{aligned} F_{mn}(P) &= v_0(P) + \sum_1^{\infty} [v_k(P) - v_{k-1}(P)], \text{ or} \\ F_{mn}(P) &= w_0(P) + \sum_1^{\infty} [w_k(P) - w_{k-1}(P)]; \end{aligned}\right\} \tag{3}$$

and by letting k be sufficiently great, we shall have uniformly

$$\left.\begin{aligned} |F_{mn}(P) - v_k(P)| &\leq \sum_{k+1}^{\infty} |v_j(P) - v_{j-1}(P)| < b^k/(1-b) < \eta/3, \text{ or} \\ |F_{mn}(P) - w_k(P)| &\leq \sum_{k+1}^{\infty} |w_j(P) - w_{j-1}(P)| < b^k/(1-b) < \eta/3. \end{aligned}\right\} \tag{4}$$

Here we shall introduce a lemma to the effect that

$$\left.\begin{aligned} |\bar{u}_{mn}(f;P) - u_{mn'}(f;P)| &\leq \max|f| \quad |\bar{u}_{mn}(1;P) - \bar{u}_{mn'}(1,P)|, \\ |u_{mn}(g;P) - u_{m'n}(g;P)| &\leq \max|1-g| \, |u_{mn}(0;P) - u_{m'n}(0;P). \end{aligned}\right\} \tag{5}$$

To prove the first part of this lemma, let $n' < n$, and let $f(P) \geq 0$. Then all we have to prove is

$$\bar{u}_{mn}(\max|f| - f;P) - \bar{u}_{mn'}(\max|f| - f;P) > 0.$$

This is easily demonstrated, however, since in general $\bar{u}_{mn}(\max|f| - f;P)$ is an increasing function of n. The extension of the theorem to non-positive f's and the demonstration of the second part of the theorem offer no difficulties.

Now let m, n, m', and n' all be so large that over the region R exterior to H and K and containing H and K

$$\left.\begin{aligned} |u_{m1}(0;P) - u_1(0;P)| &< \epsilon, \quad |u_{m'1}(0;P) - u_1(0;P)| < \epsilon, \\ |\bar{u}_{1n}(1;P) - \bar{u}_1(1;P)| &< \epsilon, \quad |\bar{u}_{1n'}(1;P) - u_1(1;P)| < \epsilon. \end{aligned}\right\} \tag{6}$$

Let $v_k'(P)$ and $w_k'(P)$ be the functions obtained, respectively, by substituting m' and n' for m and n in the definition of $v_k(P)$ and $w_k(P)$. Then by the definition of v_k and w_k, we obtain from (5)

$$\left.\begin{aligned} |v_0(P) - v_0'(P)| &< 2\epsilon, \\ |w_0(P) - w_0'(P)| &< 2\epsilon, \\ |v_1(P) - v_1'(P)| &< 4\epsilon, \\ |w_1(P) - w_1'(P)| &< 4\epsilon, \\ &\cdots\cdots \\ |v_k(P) - v_k'(P)| &< 2k\epsilon, \\ |w_k(P) - w_k'(P)| &< 2k\epsilon, \\ &\cdots\cdots \end{aligned}\right\} \tag{7}$$

By choosing ϵ so that $2k\epsilon<\eta/3$, we have from (4) and (7)

$$|F_{mn}(P)-F_{m'n'}(P)|<\eta. \tag{8}$$

In other words, $F_{mn}(P)$ converges uniformly over R to a limit which of course is harmonic. We may term this limit, which we shall write $W_{HK}(H; P)$, the weight at P of H in the system H, K. It is easy to show that $W_{HK}(H; P)+W_{KH}(K; P)=1$.

With the exception of this last sentence, all the results of this section extend at once to n dimensions. It is of course necessary to give a suitable interpretation to the exterior Dirichlet problem. As is familiar, the Dirichlet problem in three or more dimensions involves the additional condition that the harmonic function sought for shall vanish at infinity.

Another important remark is that $W_{HK}(H; P)$ is independent of the particular sequences $\{H_m\}$ and $\{K_n\}$ chosen in obtaining it. Let $\{H_m'\}$ be another sequence fulfilling the conditions laid down for $\{H_m\}$. H_1 will contain in its interior a term of $\{H_m'\}$, say H_2''. This will in turn contain in its interior a term of $\{H_m\}$, say H_3''. By continuing this process, we can construct a sequence $\{H_m''\}$, consisting alternately of terms from $\{H_m\}$ and $\{H_m'\}$. A similar remark holds of $\{K_n\}$. Clearly the value of $W_{HK}(H; P)$ formed by substituting $\{H_m''\}$ and the similarly derived sequence $\{K_n''\}$ for $\{H_m\}$ and $\{K_n\}$ respectively is at once identical with that corresponding to the sequences $\{H_m\}$ and $\{K_n\}$ and to that corresponding to the sequences $\{H_m'\}$ and $\{K_n'\}$.

§2. **The generalized Dirichlet problem.** Let R be a connected, bounded set of points in n-space, containing no point of its boundary, and let C be its boundary. Let C_1 and C_2 be two sub-sets of C, such that the lower bound of the distance from a point of C_1 to a point of C_2 is the positive number d. Then the number $W_{C_1C_2}(C_1; P)$ will exist for every point in R, and will lie between 0 and 1. If C_1 be kept fixed and C_2 be allowed to vary, $W_{C_1C_2}(C_1; P)$ will have a non-negative lower bound. This we shall term the *outer weight* of C_1 at P, and shall represent by the symbol $W(C_1; P)$. Clearly if C_1 and C_2 are defined as above,

$$W(C_1; P)+W(C_2; P)=W(C_1+C_2; P), \tag{9}$$

while if C_1 and C_2 are any two sub-sets of C,

$$W(C_1; P)+W(C_2; P)\geq W(C_1+C_2; P). \tag{10}$$

For reasons of simplicity we shall now use two-dimensional language. Let C_x be the set of all points of C for which the X-coördinate does not exceed x. Then $W(C_x; P)$ is an increasing function of x. It therefore has no discontinuities other than a denumerable set of finite jumps. It follows at once that if $\{\xi_n\}$ is a denumerable set of values of x, there is only a denumerable set of values of a for which $\{\xi_n+a\}$ contains a discontinuity of $W(C_x; P)$. Hence it is possible to determine the origin of coördinates in such a manner that no value of x that is a terminating binary number is a discontinuity of $W(C_x; P)$. Similarly, if we let C_y' be the set of all points of C for which the Y-coördinate does not exceed y, and suitably choose the vertical level of the origin of coördinates, no value of y that is a terminating binary number will be a discontinuity of $W(C_y'; P)$.

Let us choose our origin in the manner indicated above, and let x be a terminating binary number. Let D be the set of points of C with an X-coördinate not less than $x-\epsilon$ and not greater than $x+\epsilon$. Then:

$$W(C_{x-2\epsilon}; P)+W(D; P)\leq W(C_{x+\epsilon}; P). \tag{11}$$

This is an immediate result of the fact that the weight of a set of points is at least as great as that of an included set of points, and of (9). We deduce at once

$$W(D; P)\leq W(C_{x+\epsilon}; P)-W(C_{x-2\epsilon}; P), \tag{12}$$

whence we obtain

$$\lim_{\epsilon\to 0} W(D; P)=0. \tag{13}$$

This last formula holds uniformly over any closed set of points belonging to R. This follows from the fact that $W(D; P)$ is a non-negative harmonic function. Hence if C is a circle with P as center lying entirely within R, if r is the radius of C, and if Q is a point with polar coördinates (ρ, θ) with respect to P as origin, then

$$W(D;Q)=\frac{1}{2\pi}\int_0^{2\pi} W(D;Q')\,\frac{1-\rho^2/r^2}{1-2\rho\cos(\theta-\theta')/r+\rho^2/r^2}\,d\theta' \qquad (14)$$

$$\leq\frac{1}{2\pi}\int_0^{2\pi} W(D;Q')d\theta\,\frac{r+\rho}{r-\rho}=W(D;P)\,\frac{r+\rho}{r-\rho},$$

Q' being understood to be (r,θ'). It follows that (13) holds uniformly over any circle interior and concentric to C, and hence, by a process such as that used in the analytic continuation of a function, it may be proved that (13) holds uniformly over any closed set of points interior to R.

Now let $U(P)$ be any function continuous on C. There is no essential restriction on U if we assume $|U(P)|<1$. It is then possible to divide up the plane into a network of squares of vertical and horizontal lines, with opposite corners of the forms $(j/2^q,\ k/2^q)$ and $((j+1)/2^q,\ (k+1)/2^q)$, such that the oscillation of U in none of these squares exceeds θ. We may then remove from C a set D, consisting of all points in a set of vertical and horizontal strips containing all the boundary points of these squares, and in total outer weight less than ϵ. There will be left a set of sets of points of C, each consisting in all the points of C in one of the squares and not in D. We shall call this set $\{C_n\}$.

We shall now prove a lemma to the following effect: *if from each C_n we select a point P_n, and if R' is a closed set of points interior to R, then there is a function $u(P)$ harmonic over R, and such that if θ and ϵ tend to 0 in any manner whatever, then*

$$u(P)=\lim_{\theta,\epsilon\to 0}\Sigma U(P_n)W(C_n;P) \qquad (15)$$

uniformly over R'.

To prove this, let us form sets $\{C_n'\}$ and points $\{P_n'\}$ corresponding to a value ϵ' of ϵ and a value θ' of θ each less than the values corresponding to $\{C_n\}$ and $\{P_n\}$. Then the removal from the sets $\{C_n'\}$ of the points of D will leave a finite number of disconnected sets $\{C_n''\}$. Let P_n'' be a point from C_n''. We shall have

$$|\Sigma U(P_n'')W(C_n'';P)-\Sigma U(Q_n)W(C_n'';P)|\leq\theta,$$

where Q_n is the value of the P_m from the C_m containing C_n''. Moreover,

$$|\Sigma U(Q_n)W(C_n''; P)-\Sigma U(P_n)W(C_n; P)|$$
$$\leq \Sigma W(C_n; P)-\Sigma W(C_n''; P)<\epsilon.$$

This holds uniformly over R'. Hence over R' we have

$$|\Sigma U(P_n'')W(C_n''; P)-\Sigma U(P_n)W(C_n; P)|<\epsilon+\theta,$$

and likewise

$$|\Sigma U(P_n'')W(C_n''; P)-\Sigma U(P_n')W(C_n'; P)|<\epsilon+\theta;$$

whence

$$|\Sigma U(P_n')W(C_n'; P)-\Sigma U(P_n)W(C_n; P)|<2\epsilon+2\theta.$$

From this (15) may be deduced at once.

It is manifest from this last demonstration that $u(P)$ is quite independent of the particular sequences of values of C_n and P_n employed in attaining to it. We now wish to prove that $u(P)$ is the function represented by the same symbol in Theorem I. In the following discussion, Γ_ν, $v(P)$, and $u_\nu(P)$ receive the same interpretations as in the formulation of Theorem I.

Since D is of outer weight less than ϵ, it is possible to find a boundary D' for the exterior of which the Dirichlet problem is soluble, enclosing D in its interior, and such that over R' its outer weight in the system D', C is less than ϵ. Now, the outer weight of D' over R' in the system D', Γ_ν is no greater than the outer weight of D' over R' in the system D', C, provided only ν is so large that R' lies entirely within Γ_ν. To prove this, we must first notice that it is possible to find a D'' containing D' in its interior and a contour containing a closed sub-set of C, say C_1, exterior to D'', such that the Dirichlet problem is soluble over C_1, D'', and

$$|W_{D''C_1}(D''; P)-W'(D'; P)|<\eta, \tag{16}$$

over R', where W' is used as the symbol for outer weight in the system C, D'. It is also possible to find a D''' interior to D'' and a Γ_ν' exterior to C_1 and containing all points of Γ_ν exterior to or on D'', such that the Dirichlet problem is soluble within the boundary D''', Γ_ν', and such that over R'

$$|W_{D'''\Gamma_\nu'}(D'''; P)-W''(D'; P)|<\eta, \tag{17}$$

where W'' is used as the symbol for outer weight in the system Γ, D'. Now, on Γ_ν'

$$W_{D''\Gamma_\nu'}(D''';P)=0,$$

$$W_{D''C_1}(D'';P)\geq 0;$$

whereas on D''

$$W_{D'''\Gamma_\nu'}(D''';P)\leq 1,$$

$$W_{D''C_1}(D'';P)=1.$$

Since D'' and Γ_ν' entirely include R', we have

$$W_{D''C_1}(D'';P)\geq W_{D'''\Gamma_\nu'}(D''';P) \tag{18}$$

over R'. Combining (16), (17), and (18), we get

$$W'(D';P)\geq W''(D';P), \tag{19}$$

which is our desired result.

If we remove from Γ_ν all the points interior to D', and if ν is sufficiently large, Γ_ν will fall apart into a set of separated sets of points $\bar{C}_n$, each consisting of all points of Γ_ν not within D' and within the same square used in defining C_n. We now wish to establish an upper bound for

$$|W'(C_n;P)-W''(\bar{C}_n;P)|$$

for all values of P in R'.

To this end, let us remark that

$$|W'(C_n;P)-W_{C_nF}(C_n;P)|<\epsilon$$

where F is the set of all points neither in C_n nor in D. If then G and H are contours surrounding C_n and F, respectively, sufficiently closely, and the Dirichlet problem is soluble for the region exterior to H and G, while $\Phi(P)$ is the harmonic function corresponding to the boundary conditions 1 on G and 0 on H, then over R' we have

$$|\Phi(P)-W'(C_n;P)|<\epsilon. \tag{20}$$

Let G be so chosen, for some H, and let ν be so large that G contains on its interior $\bar{C}_n$. Now let H be so chosen that (20) holds, while H is entirely exterior to Γ_ν. Let us compare the functions $\Phi(P)$ and $W_{\bar{C}_n\bar{F}}(\bar{C}_n;P)$, where $\bar{F}$ consists of all points in Γ_ν but not in $\bar{C}_n$ nor within or on D'. We can approximate to $W_{\bar{C}_n\bar{F}}(\bar{C}_n;P)$ as closely as we wish by the harmonic function $\Phi_1(P)$, which is

determined by surrounding $\overline{C}_n$ by a contour G_1 and $\overline{F}$ by a contour H_1 outside of which the Dirichlet problem is soluble, and assigning boundary conditions 1 on G_1 and 0 on H_1. We may then take G_1 entirely within G and H_1 entirely without H. R' will be included by a contour formed of G, H_1, and D'. Over G and H_1 we have $\Phi \geq \Phi_1$, while neither Φ nor Φ_1 exceeds 1 on D'. It may then readily be deduced with the aid of a result similar to (19) that over R'

$$\Phi_1(P) - \Phi(P) \leq 2\epsilon. \tag{21}$$

Combining this result with (20), we can show without difficulty that

$$W_{\overline{C}_n \overline{F}}(\overline{C}_n; P) - W'(C_n; P) < 3\epsilon. \tag{22}$$

A precisely similar argument in which the rôles of C_n and F are interchanged throughout, with the correlated changes made in all the other sets of points, will lead to the result

$$W'(C_n; P) - W_{\overline{C}_n \overline{F}}(\overline{C}_n; P) < 3\epsilon. \tag{23}$$

Combining (22) and (23), we get

$$|W'(C_n; P) - W_{\overline{C}_n \overline{F}}(\overline{C}_n; P)| < 3\epsilon. \tag{24}$$

This holds uniformly over R'. Combining this with (19), we get the result

$$|W'(C_n; P) - W''(\overline{C}_n; P| < 4\epsilon \tag{25}$$

over R', since

$$W''(\overline{C}_n; P) + W''(D'; P) \geq W_{\overline{C}_n \overline{F}}(\overline{C}_n; P) \geq W''(\overline{C}_n; P)$$

There is no difficulty in proving that

$$|W'(C_n; P) - W(C_n; P)| < \epsilon, \tag{26}$$

and that, if W''' is used as the symbol for weight in the system Γ_ν, then

$$|W''(\overline{C}_n; P) - W'''(\overline{C}_n; P)| < \epsilon. \tag{27}$$

Combining (25), (26), and (27), we get over R'

$$|W(C_n; P) - W'''(\overline{C}_n; P)| < 6\epsilon. \tag{28}$$

If we now form a continuous function on Γ_ν (as we can do without difficulty in a way that we shall indicate later), which shall assume the value 1 on $\overline{C}_n$, 0 on $\overline{F}$, and intermediate values elsewhere,

then it may be shown without difficulty that the harmonic function corresponding to these boundary conditions will differ from $W'''(\overline{C}_n; P)$ by less than ϵ over R'. Let this harmonic function be $\Phi_n(P)$. We then have

$$|\Sigma U(P_n)W(C_n; P)-\Sigma U(P_n)\Phi_n(P)|<7N\epsilon \tag{29}$$

over R', where N is the number of distinct sets C_n.

Now let us compare $u_\nu(P)$ and $\Sigma U(P_n)\Phi_n(P)$. By making q sufficiently large, if there is any function $v(P)$, which we may without essential restriction suppose less than or equal to 1 in absolute value, continuous over $R+C$ and reducing to $U(P)$ on C, we can make the oscillation of $v(P)$ over the interior of each square into which $R+C$ is divided by the lines $x=j/2^q$, $y=k/2^q$ less than θ. We shall then have over $\overline{F}$, for ν sufficiently large,

$$|\Sigma U(P_n)\Phi_n(P)-u_\nu(P)|<\theta, \tag{30}$$

and over the part of Γ_ν not exterior to D'

$$|\Sigma U(P_n)\Phi_n(P)-u_\nu(P)|\leq N+1. \tag{31}$$

Now let $G(P)$ be a function defined over Γ_ν, equal to $\Sigma U(P_n)\Phi_n(P)-u_\nu(P)$ when this latter quantity is less than θ in absolute value, equal to θ when $\Sigma U(P_n)\Phi_n(P)\geq\theta$, and equal to $-\theta$ when $\Sigma U(P_n)\Phi_n(P)-u_\nu(P)\leq-\theta$. Let this give rise to a function $g(P)$ harmonic over R. Clearly the function $\Sigma U(P_n)\Phi_n(P)-u_\nu(P)-G(P)$ is zero over $\overline{F}$, and nowhere exceeds $N+1+\theta$ in absolute value. Hence by comparison of this latter function with $W'''(D'; P)$, we get the result

$$|\Sigma U(P_n)\Phi_n(P)-u_\nu(P)|\leq\theta+(N+1+\theta)\epsilon \tag{32}$$

uniformly over R' Combining this result with (15), we get

$$|u(P)-u_\nu(P)|<\theta+(N+1+\theta)\epsilon+\sigma(\epsilon, \theta), \tag{33}$$

where σ vanishes as θ and ϵ tend to zero independently. Relation (33) will hold for all sufficiently great values of ν.

Now let ϵ and θ be chosen so small that $\theta<\eta/3$ and $\sigma(\epsilon, \theta)<\eta/3$. N will be determined by θ. We may then impose the further condition on ϵ that $(N+1+\theta)\epsilon<\eta/3$. We shall then have

$$|u(P)-u_\nu(P)|<\eta \tag{34}$$

over R' for all sufficiently large ν. That is,

$$u(P) = \lim_{\nu \to \infty} u_\nu(P) \tag{35}$$

uniformly over R', and Theorem I is proved in the two-dimensional case, provided only that we show the existence of boundaries Γ_ν and of an interpolation function $v(P)$.

The boundaries Γ_ν can be constructed with the greatest ease. All we need to do is to form the meshwork of lines $x = j/2^\nu$, $y = k/2^\nu$ in the plane, and to take as our Γ_ν that part of the boundary of the set of all points on squares of the net containing C which lies in R. We may then take for the value of $v(P)$ at each of the nodes of this meshwork lying on Γ_ν the value of $U(P)$ at some point on a mesh on which the node lies. Between the nodes, we may define $v(P)$ on Γ_ν by linear interpolation. The Dirichlet problem is soluble over the region bounded exteriorly by $\Gamma_{\nu+1}$ and interiorly by Γ_ν. Let $v(P)$ be defined over this region as the harmonic function corresponding to the assigned values of $v(P)$ on the periphery. There will be no difficulty in showing[6] that the $v(P)$ thus defined is continuous over $R+C$, if we give it the values $U(P)$ on C. It is to be noted that the $v(P)$ thus defined assumes its extreme values on C.

We have already obtained $u(P)$ in a manner independent of $v(P)$ and the boundaries Γ_ν. We have therefore given a complete demonstration of the validity of Theorem I in the interior two-dimensional case. However, neither the extension to n dimensions nor the obvious verbal alterations which are necessary

[6] Let P be a point on C and let Q be a point of R at a distance not exceeding ϵ from P. Draw a circle S of radius 2ϵ with P as center. Let Q lie on the zone between Γ_ν and $\Gamma_{\nu+1}$. Let the angular measure of the points in this zone and on the periphery of S be M_ν. Since $\Sigma M_\nu \leq 2\pi$, it follows that $\lim M_\nu = 0$. Let B_1 be the upper bound of $|U(P) - U(P_1)|$ if $\overline{PP_1} \leq 2 + 1/2^{\nu-1}$ and let B_2 be the upper bound of $|U(P) - U(P_2)|$ for any P_2 on C. Then $|U(P) - v(Q)|$ is not greater than the potential at Q due to boundary values on S not exceeding B_2 anywhere and not exceeding B_1 except on a set of measure M_ν. That is,

$$|U(P) - v(Q)| \leq 3M_\nu B_2 + B_1.$$

It follows at once that $\lim_{Q \to P} v(Q) = U(P)$.

With regard to the entire theorem, cf. Lebesgue, *Sur le problème de Dirichlet*, Rendiconti di Palermo, v. 24 (1907), pp. 371–402.

to make Theorem I apply to the exterior Dirichlet problem changes the demonstration of Theorem I in a single essential respect.

The generalized solution which we have obtained for the Dirichlet problem has the following properties, which are merely generalizations of the familiar properties of the solution of the ordinary Dirichlet problem and which may be proved without any difficulty:

(1) The generalized solution of the Dirichlet problem lies between the upper and lower bounds of the given boundary values;

(2) The generalized solution of the Dirichlet problem is at every point an additive functional of the boundary values;

(3) The generalized solution of the Dirichlet problem coincides with the solution of the ordinary Dirichlet problem, if the latter exists;

(4) If the boundary C of R breaks up into two separated pieces, C_1 and C_2, and if $v(P)$ and Γ_n are defined as in the formulation of Theorem I, then if Γ_n' represents that part of Γ_n approximating to C_1, and the Dirichlet problem is soluble over the region within $\Gamma_n'+C_2$, then the function $u(P)$ of Theorem I is the uniform limit over any set of points R' closed and interior to R of the harmonic function corresponding to the boundary conditions $v(P)$ over Γ_n' and $U(P)$ over C_2. It is not indeed necessary that the Dirichlet problem be soluble over the region within $\Gamma_n'+C_2$ for boundary conditions other than these special ones.

It results immediately from (4) that if the generalized Dirichlet problem has the solution $u(P)$ for the boundary condition $U(P)$ on C, and C' is a boundary entirely interior to C, then $u(P)$ will be the solution of the generalized Dirichlet problem corresponding to boundary conditions $U(P)$ on C and $u(P)$ on C'.

These results will be assumed without explicit reference in the sequel.[7]

[7] The generalization of the Dirichlet problem in the present paper consists in a generalization of the boundary permissible, the boundary conditions remaining continuous. In a paper forthcoming in the *Trans. Am. Math. Soc.*, the author of this paper shows how for a region for which the Dirichlet problem is soluble for continuous boundary conditions, a generalized Dirichlet problem is soluble for a wide class of discontinuous boundary conditions. The two types of generalization of the Dirichlet problem thus are in two

§3. **The distribution of Charge on an isolated conductor.** It is a familiar physical fact that if an electric charge is released on an isolated conductor, it will spread over the surface of the conductor without gain or loss until the latter reaches a state of equilibrium. The potential caused by this charge will then be constant over the surface of the conductor. It accordingly becomes a matter of interest to ascertain whether the general exterior boundary value problem of Laplace's equation for boundary values 1 can be solved by the potential due to a surface distribution of charge. Of course this problem in its unmodified form is only of interest in three or more dimensions. For purposes of definiteness we shall deal with the three-dimensional case.

As is well known, if S is any closed surface of finite area possessing at every point a normal, and if $U(P)$ is any function harmonic within and on S, we have

$$\iint_S \partial U/\partial n \; dS = 0, \tag{36}$$

where $\partial U/\partial n$ is the derivative of U along the internal normal. It is accordingly easy to prove that if $U(P)$ is a harmonic function over the whole of space, except for sets of points $A_1, A_2, \ldots, A_n$, and if S is a closed boundary with a normal everywhere and over which U is harmonic, then if A_1 is within S while all other A's are exterior to S, $\iint_S \partial U/\partial n \; dS$ is independent of S, and is determined by U and A_1 alone. We shall call $1/4\pi \iint_S \partial U/\partial n \; dS$ the *charge* on A_1.

Now let H and K be two boundaries entirely exterior to one another, and let $v(P)$ be the harmonic function which corresponds in the sense of the last section to the boundary values $V(P)$ on H and K and 0 at ∞. Let $v_1(P)$ correspond to the boundary values $V(P)$ on H and 0 at ∞, K not being considered as part of the boundary. Let $v_2(P)$ correspond to the boundary values

essentially distinct directions. The methods of the *Transactions* paper, however, are as directly applicable to regions for which the Dirichlet problem is only soluble for continuous boundary conditions in the extended sense of this paper as to the cases there treated. Indeed all of that paper, except for so much as relates to the continuity of the solution of the generalized Dirichlet problem near the boundary, remains unchanged if the generalized Dirichlet problem in the sense of this paper is substituted for the ordinary Dirichlet problem.

$V(P)-v_1(P)$ on K and 0 at ∞, H not being regarded as part of the boundary. In general, let v_{2n+1} correspond to the boundary values $V(P)-\sum_1^{2n} v_k(P)$ on K and 0 at ∞, while V_{2n} corresponds to boundary values $V(P)-\sum_1^{2n-1} v_k(P)$ on H. It may be shown with the aid of considerations in no wise different from those familiar in the ordinary instances of the alternating process that

$$v(P)=\sum_1^{\infty} v_n(P) \tag{37}$$

uniformly over the entire exterior of H and K, as will result from the fact that there is a convergent geometric progression dominant to (37). It hence follows that the series

$$\left.\begin{aligned} v'(P)&=\Sigma v_{2n}(P)\\ v''(P)&=\Sigma v_{2n+1}(P)\end{aligned}\right\} \tag{38}$$

converge uniformly over the region exterior to K and that exterior to H, respectively, and are harmonic over these regions. They vanish at ∞.

It is not possible to break up $v(P)$ in any other way into two functions vanishing at infinity, one harmonic outside H, and the other harmonic outside K. For suppose that

$$v(P)=u'(P)+u''(P),$$

where u' is harmonic outside H and u'' is harmonic outside K. Then outside $H+K$

$$u'(P)-v'(P)+u''(P)-v''(P)=0.$$

It follows that $u'(P)-v'(P)$ and $v''(P)-u''(P)$ are part of the same harmonic function, which thus has no singularities in the finite plane and vanishes at infinity. It is therefore identically zero.

We may conclude at once that if $U(P)$ is the harmonic function corresponding to boundary values zero at infinity and 1 on the boundary made up of the non-intersecting closed portions A_1, A_2, . . . , A_n, then there are uniquely determined functions U_1, U_2, . . . , U_n, such that $U(P)=\Sigma U_k(P)$, while $U_k(P)$ is harmonic outside A_k and vanishes at infinity. We may verify the fact that the charge on A_k corresponding to the potential

U_k is the same as the charge on A_k corresponding to the potential U. We may indeed regard U_k as that part of U generated by charges on A_k.

U_k is positive or 0 on A_k. To show this, let us call A_k H and the sum total of the other A's K, and let us take 1 for $V(P)$ in the process leading to (37). Then $v''(P) = U_k(P)$. We may readily verify that

$$\left.\begin{aligned} &v_1 \text{ corresponds to a boundary condition 1 on } H.\\ &v_2 \text{ corresponds to a boundary condition } 1-v_1 \text{ on } K,\\ &v_3 \text{ corresponds to a boundary condition } -v_2 \text{ on } H,\\ &v_4 \text{ corresponds to a boundary condition } -v_3 \text{ on } K,\\ &\quad\cdots\cdots\cdots\cdots\cdots\cdots\\ &v_{2n} \text{ corresponds to a boundary condition } -v_{2n-1} \text{ on } K,\\ &v_{2n+1} \text{ corresponds to a boundary condition } -v_{2n} \text{ on } H. \end{aligned}\right\} \quad (39)$$

Hence

$$\left.\begin{aligned} &v_1+v_3+\ldots+v_{2n+1} \text{ corresponds to } 1-v_2-\ldots-v_{2n} \text{ on } H,\\ &v_2+v_4+\ldots+v_{2n} \text{ corresponds to } 1-v_1-\ldots-v_{2n-1} \text{ on } K. \end{aligned}\right\} \quad (40)$$

It is hence easy to prove by mathematical induction that the partial sums of the odd v's and also the partial sums of the even v's always lie between 0 and 1. Hence both v' and v'' are non-negative. Since U_k thus corresponds to boundary values non-negative on A_k, we may conclude at once that the charge on A_k corresponding to U is likewise non-negative, as a consideration of the gradient of U_k along any large sphere containing A_k will show.

Now let C be the boundary of an open region R extending to infinity, let C be bounded, and let U be the harmonic function corresponding to the boundary values 1 on C and 0 at infinity. Let A be a sub-set of C, and let B be a sub-set sharing no point or limit-point with A. Let U' be the harmonic function corresponding to the boundary conditions 1 on $A+B$ and 0 at infinity. Let σ' be the charge on A corresponding to U'. Then as B increases in such a manner as to make $A+B$ approach C, σ' will approach a definite positive limit σ, which we shall

term the outer charge on A corresponding to the potential distribution U.

To show this, let B' be a set of points containing B as a sub-set, but forming part of C and sharing no point or limit-point with A. Let U'' be the harmonic function corresponding to the boundary values 1 on $A+B'$ and 0 at infinity. Then $U''-U'$ will correspond to boundary values non-negative on B', 0 on A, and 0 at infinity. In the alternating process leading to (37) let H be A, let K be B', and let $V(P)$ be $U''-U'$ on B' and 0 on A. Then

$$\left.\begin{aligned} &v_1 \text{ corresponds to a boundary condition 0 on } H, \\ &\quad \text{and vanishes,} \\ &v_2 \text{ corresponds to a boundary condition } V \text{ on } K, \\ &v_3 \text{ corresponds to a boundary condition } -v_2 \text{ on } H, \\ &v_4 \text{ corresponds to a boundary condition } -v_3 \text{ on } K, \\ &\quad \cdots\cdots\cdots\cdots\cdots\cdots \\ &v_{2n} \text{ corresponds to a boundary condition } -v_{2n-1} \text{ on } K, \\ &v_{2n+1} \text{ corresponds to a boundary condition } -v_{2n} \text{ on } H. \end{aligned}\right\} \quad (41)$$

Hence

$$\left.\begin{aligned} &v_1+v_3+\ldots+v_{2n+1} \text{ corresponds to } -v_2-\ldots-v_{2n} \text{ on } H, \\ &v_2+v_4+\ldots+v_{2n} \text{ corresponds to } V-v_3-\ldots-v_{2n-1} \text{ on } K. \end{aligned}\right\} \quad (42)$$

We may then conclude by mathematical induction that the sums of the odd v's are non-positive on H, and also that the sums of the even v's exceed or equal V on K. Hence $v''(P)$ is negative, and σ'', the charge on A corresponding to U'', is less than σ'.

It follows that if $B_1, B_2, \ldots$ is a sequence of B's, each including its predecessor, while $\sigma_1, \sigma_2, \ldots$ are the corresponding σ's, then the sequence of σ's is a decreasing sequence of positive numbers and hence has a limit. If finally every point of C not a point or limit point of A is in some B_n, it is easy to show that the value obtained for $\sigma = \lim \sigma_n$ is unique and independent of the sequence of B_n's.

A theorem which is important, and which can be verified without difficulty is the following: if A and B are two sub-sets of C,

the outer charge on $A+B$ does not exceed the outer charge on A, added to the outer charge on B. This is an immediate corollary of the fact that if D is a set of points sharing no point or limit-point with $A+B$, the charge on A corresponding to a potential of 1 on $A+B+D$ added to that on B is at least that on $A+B$.

Let C be defined as before. Let $C(x)$ be that part of C made up of those points with an X-coördinate not exceeding x. If $\sigma(x)$ is the outer charge on $C(x)$ corresponding to a potential 1 on C, it is an increasing function of x. As such it has only a denumerable set of discontinuities. As in Section 2, it is then possible to choose the origin of coördinates in such a manner that $\sigma(x)$ will have no discontinuities for any terminating binary x. Making a similar choice of the origin with regard to y and z, we find it possible to enclose the planes $x=i/2^n$, $y=j/2^n$, $z=k/2^n$ in so far as they consist of points of C in the interior of a closed set of points of arbitrarily small outer charge — say of outer charge less than ϵ. We may simultaneously make the outer weight of this set $<\epsilon$ over a closed set R' exterior to C.

The portion of C that remains will be divided up into a number of separated sets of points C_m, each of maximum linear dimensions not exceeding $1/2^{n-1}$. Let us investigate the distribution of charge on these sets corresponding to a boundary potential of 1 on ΣC_m. We shall find on each C_m a charge σ_m, and the potential due to the charge on C_m alone will be positive over all C_m. It will hence be positive over a sphere of radius $1/2^n$ containing C_m. It will of course not exceed 1. Let us expand the potential on the exterior of this sphere S in the Poisson integral.

$$U_m(P)=\frac{1}{4\pi a}\int\int_S \frac{U_m(Q)(r^2-a^2)dS}{[r^2+a^2-2ar\cos(a,r)]^{3/2}}, \tag{43}$$

where P is a point on the exterior of S, Q is a point on the element dS of S, P has the polar coördinates (r, θ, ϕ), Q the coördinates (a, θ', ϕ'), U_m represents the potential on the surface of the sphere, and (a, r) is the angle between OQ and OP, O being the origin and the center of S. Now,

$$\frac{r-a)}{(r+a)^2}\leq\frac{r^2-a^2}{[r^2+a^2-2ar\cos(a,r)]^{3/2}}\leq\frac{r+a}{(r-a)^2}, \tag{44}$$

from which it follows that

$$\left|\frac{1}{r}-\frac{r^2-a^2}{[r^2+a^2-2ar\cos(a,r)]^{3/2}}\right| \leq \frac{4a}{(r-a)^2}. \tag{45}$$

Therefore

$$\left|u_m(P)-\frac{1}{4\pi a}\int\int_S \frac{U_m(Q)dS}{r}\right| \leq \frac{4a}{(r-a)^2}\frac{1}{4\pi a}\int\int_S U_m(Q)dS. \tag{46}$$

However, $1/4\pi a \int\int_S U_m(Q)dS$ is the value of the charge on the sphere, so that we see that the change in potential at P which would result from concentrating the entire charge of C_m at O cannot exceed the fraction $\frac{4ar}{(r-a)^2}$ of that potential.

Now let R' be a set of points at a minimum distance of d from C. If $w(P)$ is the harmonic function corresponding to a boundary potential of 1 on ΣC_m, and if n is so large that

$$\frac{4ad}{(d-a)^2} < \eta, \tag{47}$$

then

$$|w(P)-\Sigma\sigma_m/r_m| < 2\eta, \tag{48}$$

where r_m is the distance from P to some point Q_m in C_m.

We know that by making ϵ sufficiently small we can make $w(P)$ uniformly approach over R' the potential corresponding to boundary conditions 1 over C. This we have substantially shown in Section 2. Let us now investigate the change in $\Sigma\sigma_m/r_m$ due to the decrease of the region D within which we have enclosed our planes $x=i/2^n$, $y=j/2^n$, $z=k/2^n$. We know that the outer charge on C_m+D corresponding to a potential 1 on C is at least as great as σ_m. On the other hand, the argument of (41) and (42) has shown that the outer charge on C_m corresponding to a potential 1 on C does not exceed σ_m. From this and the theorem which we have given concerning the outer charge on the logical sum of two sets, it results at once that if C_m' is the set of all points of C in the cube bounded by planes $x=i/2^n$, $y=j/2^n$, $z=k/2^n$ containing C_m, then the outer charge on C_m' corresponding to a potential 1 on C does not differ from σ_m by more than ϵ. If then this outer

charge be σ_m', and if there be N cubes containing C_m's, we shall have

$$|U(P)-\Sigma\sigma_m'/r_m|<2\eta+N\epsilon+\phi(\epsilon), \tag{49}$$

where $U(P)$ corresponds to boundary values 1 on C and O at infinity, and $\phi(x)$ vanishes with x. Since ϵ is not involved in the determination of the C_m's', (49) reduces to

$$|U(P)-\Sigma\sigma_m'/r_m|\leq 2\eta. \tag{50}$$

Let $M(P)$ stand for the outer charge on the set of all points Q of C such that

$$x<x',\ y<y',\quad z<z'; \tag{51}$$

(x, y, z) and (x', y', z') being the coördinates of Q and P, respectively. The $M(P)$ is an increasing function of all the coördinates of P, and is continuous whenever the coördinates of P are terminating binary numbers. In the language of Daniell[8] it is of positive type. While Daniell is not very explicit as to his definition of a multiple Stieltjes integral, it is clear that since

$$\sigma_m'=\Delta^3_{ijk}M \tag{52}$$

in his notation, (x_{1i}, x_{2j}, x_{3k}) and $(x_{1,i+1}, x_{2,j+1}, x_{3,k+1})$ being opposite vertices of the cube containing C_m, we shall have as a consequence of (50) and (47)

$$U(P)=\int_{a_1}^{b_1}\int_{a_2}^{b_2}\int_{a_3}^{b_3}(PQ)^{-1}dM(Q). \tag{53}$$

The rest of Theorem II is obvious in the three-dimensional case. In spaces of dimensionality higher than 3, the proof requires no substantial modification. In space of two dimensions, we meet the complication that the potential due to an isolated unit point charge, instead of vanishing at infinity, has a negative logarithmic infinity. The specification of the exterior Dirichlet problem is accordingly a more difficult matter. We cannot require that the potential shall remain finite at infinity without rendering our entire problem nugatory, as then the exterior Dirichlet problem corresponding to boundary conditions 1 will be a constant. On

[8] P. J. Daniell, *Functions of Limited Variation in an Infinite Number of Dimensions*, Annals of Mathematics, vol. 21 (1919–1920), p. 31.

the other hand, there are an infinity of harmonic functions corresponding to the value 1 on the boundary and with a logarithmic infinity at infinity.

The direct two-dimensional analogue of Theorem II reads: *let R be any bounded set of points in the plane, entirely within a rectangle with sides parallel to the axes and with opposite corners (a_1, a_2) and (b_1, b_2). Let $u(P)$ be a harmonic function with no singularities in the entire plane except on R, except that at infinity there is a logarithmic singularity such that*

$$\lim_{\rho \to \infty} \frac{-\log r}{u(P)} = 1,$$

r being the distance of P from the origin. Let $u(P)$ correspond to boundary values 1 over R. Then there will be a function $M(P)$ representing the charge on that portion of the boundary of R consisting of points with both coördinates less than those of P. $M(P)$ will be an increasing function of the coördinates of P. Over every point Q exterior to R we shall have

$$u(Q) = \left\{ \int_{a_1}^{b_1} \int_{a_2}^{b_2} (-\log \overline{PQ}) dM(P) \right\} \Big/ \int_{a_1}^{b_1} \int_{a_2}^{b_2} dM(P). \tag{53}$$

Moreover, if the rectangle with opposite vertices (x_1, x_2) and (y_1, y_2) is entirely exterior to R, then

$$\int_{x_1}^{y_1} \int_{x_2}^{y_2} dM(P) = 0. \tag{54}$$

In order to make the formulation of this theorem definite, we must determine what we mean by the harmonic function behaving like $-\log r$ at infinity and corresponding to boundary values 1 over R. We shall mean by this the harmonic function $w(P)$ $-\log r$, where $w(P)$ is finite at infinity, and corresponds in the sense of Theorem I to the boundary conditions $1+\log r$ on the exterior of R.

Let $v(P)$ be the harmonic function corresponding to the boundary conditions 1 on R and $-\log X$ on Σ, a circle with radius X about the origin. Let R' be a bounded set of points exterior to R. It may then be shown without difficulty that

$$u(P) = \lim_{x \to \infty} v(P) \tag{55}$$

uniformly over R'. Moreover, $t(P)=v(P)+\log X$ is a harmonic function corresponding to boundary conditions $1+\log X$ on R and O on Σ.

If R breaks up into the separated portions $A_1, A_2, \ldots, A_n$, it is possible to prove just as in the earlier part of this section that $t(P)$ may be decomposed in a unique fashion into harmonic functions $t_k(P)$, where $t_k(P)$ corresponds to a zero boundary value on Σ and has no other singularity except on A_k. The charge σ_k' on A_k corresponding to t, represented by the integral around A_k of the normal derivative of t, will be the same as the charge on A_k corresponding to t_k, and will approach a limit σ_k as X becomes infinite. As before, it may be shown that t_k corresponds to positive boundary conditions on the exterior of A_k. As before, moreover, by substituting for the exterior Poisson expansion the expansion for the solution of the Dirichlet problem over the region included between two circles, it may be shown that the difference between t_k and the potential corresponding to a charge σ_k' on a point of A_k and vanishing on Σ approaches O in comparison with σ_k uniformly over R' for any system of A_k's as the maximum dimension of any A_k approaches zero. Let the potential function corresponding to a unit charge on the point Q_m of A_m and O on Σ be M_m. Then

$$\left| t(P)-\Sigma\sigma_m'M_m \right|$$

admits a uniformly dominant expression vanishing with the greatest dimension of any A_m, over all R'. The same is therefore true of

$$\left| v(P)-\Sigma\sigma_m'\left(M_m-\frac{\log X}{\Sigma\sigma_m'}\right) \right|.$$

Now, $M_m-\dfrac{\log X}{\Sigma\sigma_m'}$ is the harmonic function corresponding to a charge of σ_m' at Q_m and assuming the boundary value $-\dfrac{\log X}{\Sigma\sigma_m'}$ on Σ. It may be shown with ease by an actual evaluation that

$$-\sigma_m \log \overline{PQ}_m = \lim\left(M_m-\frac{\log X}{\Sigma\sigma_m'}\right). \qquad (56)$$

Hence over R' $\quad |u(P)+\Sigma\sigma_m \log r_m|$

admits a uniformly dominant expression vanishing with the greatest dimension of any A_m.

From this point on the proof of our two-dimensional theorem follows lines in no essential wise different from those already used in proving Theorem II.

§4. **Capacity, and a sufficient condition for the solubility of the Dirichlet problem.** The capacity of a region R in space of more than two dimensions is the charge corresponding to a potential 1 on R and 0 at infinity. The capacity of a region R in two dimensions is the charge corresponding to a potential 1 on R, which is the boundary-value of a harmonic function behaving like $-\log r$ at infinity. In both cases, the total charge will be represented by

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} dM(P)$$

in the terminology of Theorem II or its two-dimensional analogue. It results from these theorems that every region that is bounded will have a capacity, and that the potential corresponding to boundary conditions c on a region R of capacity C at any point P is that due to a charge cC at a distance from P between that of the nearest and the most remote point of R.

Let us now proceed to the demonstration of Theorem III. While the proof is in no essential manner dependent on the dimensionality of our space, we shall for purposes of definiteness treat the three-dimensional case. Using the terminology of the enunciation of our theorem, let us consider the potential within S corresponding to a potential 1 on the part of C within S. This cannot be less than $C_1/2r=C_1/2C_2$. On S', a sphere about Q with radius ar, the same harmonic function cannot exceed $C_1/(a-1)r=C_1/(a-1)C_2$. It may be concluded without difficulty that there is a harmonic function not exceeding 1 on the part of C within S' for its boundary conditions, nor 0 on the part of S' within or on C, which is at least

$$\frac{C_1(a-3)}{C_2 2(a-1)}$$

within S. If $a=4$, this fraction will be $C_1/6C_2$.

On the other hand, the potential in some neighborhood of Q corresponding to a potential at least 1 on the part of C within S and at least 0 on the part of C without or on S will be at least the potential corresponding to a potential at least 1 on that part of S without C and 0 on that part of S within C, as may be shown by a consideration of the functions $u_n(P)$ of Theorem I. Hence a consideration of the Poisson integral shows that there is a sphere S'' about Q over which the potential function in question will be at least $1-L-\epsilon$, for any positive given ϵ.

Let us now investigate the potential function corresponding to a set of continuous boundary values on C always positive, nowhere exceeding 1, and 1 on the part of C within a certain sphere Σ about the point Q of C. It follows from the hypothesis of Theorem II and the remarks we have just made that there is a sphere Σ_1 about Q within which this harmonic function, which we shall call $W(P)$, is at least either $1-N-\epsilon$ or $M/6$, whichever is less. Let us call the lesser of these quantities d. $W(P)-d$ is non-negative in and on the boundary of Σ_1, and corresponds to boundary values at least $1-d$ on the part of C within Σ_1. An exactly similar argument to that which we have already carried through will show that there is a sphere Σ_2 about Q over which

$$W(P) \geq d+d(1-d) = 1-(1-d)^2. \tag{57}$$

It is in the same way possible to show that there is a sphere Σ_n about Q over which

$$W(P) \geq 1-(1-d)^n. \tag{58}$$

Hence $W(P)$ approaches the value 1 as p approaches Q by any route in R.

It may be observed that it is not strictly necessary for this that d be absolutely independent of r. Provided the denumerable set of values of r mentioned in the formulation of Theorem III diminish sufficiently rapidly, if d_k is the d corresponding to the k-th r, and $\prod(1-d_k)$ diverges to 0, $W(P)$ will still tend to 1 as P tends to Q.

It is at once possible to proceed to Theorem III in all its generality. In the first place, it is possible to describe about Q a sphere so small that the oscillation of U within it is less than ϵ. Hence

a modification of U and consequently of u less than ϵ will render U constant over the part of C within this sphere. This modified $u(P)$ will lie between two bounds which our demonstration has already proved to be continuous at Q, with $U(Q)$ for their limit. Hence the oscillation of u in the neighborhood of Q cannot exceed ϵ, and vanishes.

§5. **The im Kleinen character of the Dirichlet problem.** We shall now proceed to the proof of Theorem IV. Let U be a non-negative function vanishing at Q and assuming the values 1 on the part of C exterior to a sphere Σ within Q as center and interior to S. The removal of the part of C outside Σ will not increase u, and will leave it continuous at Q. It follows that the function corresponding to *any* continuous boundary condition on the remaining part of C is continuous at Q.

By first removing a small portion of C' near Q, and then employing considerations similar to those in the paragraph preceding (15), it may be shown that we can choose a sphere Σ' interior to S so that the outer weight with respect to C' of that part of C' in the neighborhood of Σ' can be made arbitrarily small. If we discard the portions of C' in progressively smaller neighborhoods of Σ', we shall therefore replace $v(P)$ by functions of a sequence $v_n(P)$ converging uniformly to v over the part of R near Q. Now, $v_n(P)$ may be obtained by the alternating process as a uniformly convergent series of functions which are harmonic and either have continuous boundary conditions on the part of C' within Σ', and are hence by the last paragraph continuous at Q if they are given their nominal boundary values on C', or are functions which are continuous at Q because Q is interior to their boundary. Hence every v_n, and consequently v, is continuously convergent to $V(Q)$ as we approach Q by a path interior to R'.

This theorem may be summed up in the statement that the property of a boundary that the solution of the generalized Dirichlet problem be continuous in the neighborhood of a given point, is an *im Kleinen* property of that portion of the boundary in the neighborhood of the point in question.

EFFECTS OF WOOD VANES IN PHONOGRAPH HORNS

By W. R. Barss, J. A. Abbott and H. M. Chatto

Interest in the effect of wood vanes in phonograph horns was first aroused on noticing what seemed to be a vastly superior performance of a phonograph having a special built-in horn of four-sided pyramidal shape and containing several wood vanes running from front of horn nearly to the small end where the tone arm of the phonograph was attached. Scratchy records were rendered in a much more pleasing manner, indicating a strong interfering out effect of the vanes. When orchestral selections were played sometimes two or three additional instruments could be plainly distinguished over the number on the usual phonograph. Certain tones seemed to be greatly reinforced, indicating an increased resonance.

Although personal observations of such a sort do not offer dependable information for scientific use, they served to convince the hearers that the presence of wood vanes improved the quality of tone, and paved the way for this investigation.

The object of this investigation was, therefore, to compare the intensity of tone reproduced by a horn having wood vanes, with that reproduced by a horn without vanes, this comparison to be made over a wide range of frequencies.

Description of Apparatus

Part of the apparatus was used to produce a tone of frequency variable over the usual audible range, and part to measure the relative intensities of the same sound when put through first one horn and then the other.

The part used to produce the tone is shown in diagram in Fig. 1. It consisted of a sonometer having a steel piano wire *WW* under tension, mounted on a resonance box *BB*. The length of string under vibration and therefore the frequency of the tone produced could be varied by means of the movable fret *F*. The string was kept vibrating by means of an air jet *J*. A microphone *M*, of special construction for reproducing phonograph records in con-

junction with a loud speaker, was placed so as to just touch the vibrating wire, and was connected through a transformer T, amplifying system A, double pole double throw switch S, to loud speakers L and L_1 mounted at the back of horns H and H_1. By means

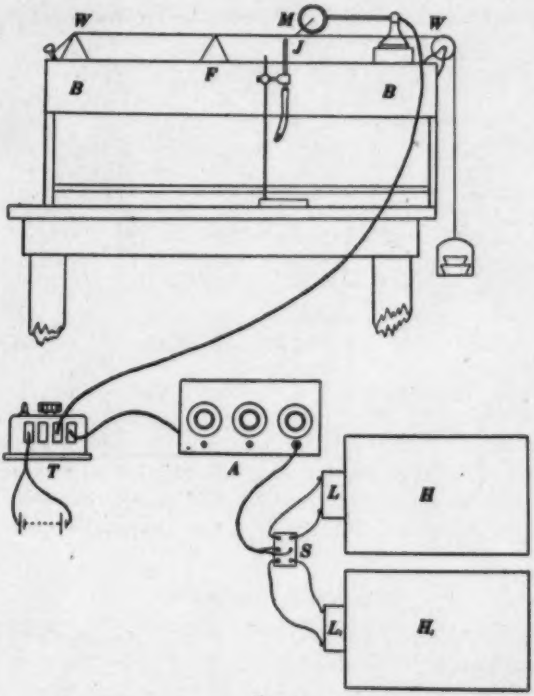


Fig 1

of S the sound could be sent into either horn. H is the horn containing vanes. H_1 is the horn without vanes. Details of the amplifying system are shown in Fig. 4. The transformer T does not play much of a part except that, by trial, it seemed to give the proper grid excitation for the first tube of the audio amplifier.

A wire .0105 inch in diameter, under a tension of 3 kilograms, was used throughout the experiments.

The rest of the apparatus, see Fig. 1a, which was used to measure the relative intensity of the sound, consists of a phonelescope[1] P', arc light A, condensing lens L, water box B and screen S. The phonelescope consists chiefly of a small mirror mounted on a

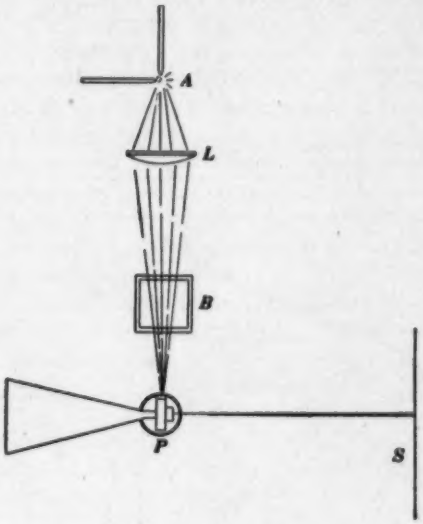


Fig 1a.

diaphragm and a small metal horn for focusing the sound on the diaphragm. The other elements are used to produce and reflect on a screen a beam of light moving according to the motion of the diaphragm. By measuring the amplitudes of the spot of light on the screen, a measure of the relative intensities of sounds incident on the diaphragm is obtained.

[1] See Dorsey Bulletin for detailed description.

A sectional diagram of the horn with vanes is shown in Fig. 2. The arrangement of vanes is such that there are no duplications of shapes anywhere in the passages between vanes. This is done with the hope of securing at least one resonant portion of the horn for all frequencies and to obtain whatever effect there may be from wood vibrating sympathetically. The fastening of the

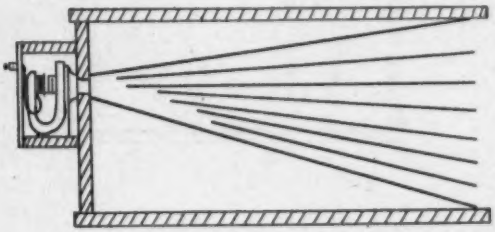


Fig 2.

vanes is an important feature in the construction of the horn. It was necessary to fasten them with glue and brads and to subject them to considerable end compression to avoid the possibility of any loose material which would cause an undesirable noise. The horn without vanes is of the same material, dimensions and construction. Mahogany was used for both horns and vanes.

Procedure and Results

The frequencies, as given by various lengths of the sonometer wire, were calibrated by means of a series of standard resonators, the range of frequencies under observation being from 120 to 960 vibrations per second. All tones of the major scale for three octaves were investigated for seven positions of the phonelescope with respect to each phonograph horn. These positions are indicated in Fig. 3, and the number associated with each position will appear on the data sheets for identification of the position concerned.

It was found, by trial, that the natural period of the phonelescope

diaphragm was outside the range of frequencies employed, so that points of high intensity from this cause were eliminated.

The next step was to calibrate one loud speaker in terms of the

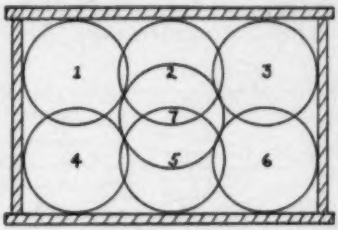


Fig 3.

other. This was done by allowing the sound to pass directly from the loud speaker to the phonelescope without passing through the intermediate wood horns. Assuming the source to remain con-

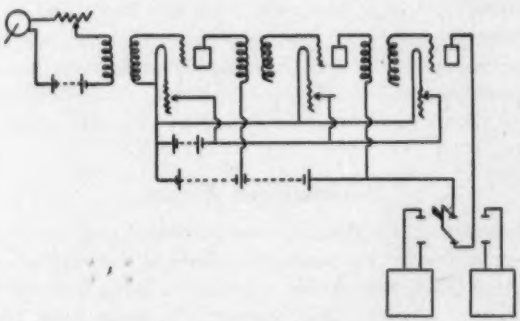


Fig 4.

stant in intensity, the relative displacements of the phonelescope diaphragm would be a relative measure of the effects of the two loud speakers and the amplifying system.

TABLE 1

Frequency	*Ratio No. 2 to No. 1*
960	2.11
900	1.56
800	1.13
720	2.67
640	3.20
600	1.50
540	2.00
480	1.67
450	2.00
400	3.00
360	2.00
320	1.29
300	1.67
270	2.50
240	2.00
225	1.50
200	2.50
180	1.57
160	1.75
150	1.80
135	1.67
120	1.67

Calibration data are shown in Table 1, where the ratio of deflection of No. 2 loud speaker to that of No. 1 loud speaker is shown against corresponding frequency.

The phonograph horns were then introduced and seven sets of readings were taken corresponding to the seven positions of the phonelescope as indicated in Fig. 3. Table 2 shows the results for position No. 1. Column 2 of this table gives the phonelescope deflection for horn No. 1 without vanes, column 3 is obtained by multiplying data in column 1 by the ratios shown in Table 1, column 2. Column 4 gives the phonelescope deflections for horn No. 2, with vanes, and column 5 is the ratio of column 4 to column 3.

Detailed data for the other positions of the phonelescope are omitted. Table 3 shows the ratios of No. 2 to No. 1 for all seven positions against corresponding frequencies.

TABLE 2

Frequency	*No.* 1	*No.* 1 *Corrected*	*No.* 2	*Ratio No.* 2 *to No.* 1
960	0.66	1.39	1.50	1.08
900	1.33	1.76	2.13	1.21
800	1.13	1.28	2.19	1.71
720	1.56	4.16	2.88	0.69
640	1.94	6.21	3.75	0.60
600	0.13	0.19	1.25	6.58
540	0.13	0.26	0.31	1.19
480	0.63	1.05	1.38	1.31
450	0.06	0.12	0.31	2.58
400	0.13	0.39	0.69	1.77
360	0.13	0.26	0.50	1.92
320	0.44	0.57	3.80	6.67
300	0.13	0.24	0.94	3.91
270	0.06	0.15	0.30	2.53
240	0.31	0.62	1.00	1.61
225	0.44	0.66	0.88	1.33
200	1.38	3.45	3.25	0.94
180	0.75	1.18	1.50	1.27
160	0.75	1.31	2.50	1.91
150	1.00	1.80	2.00	1.11
135	0.88	1.47	1.50	1.02
120	1.00	1.67	1.38	0.83

TABLE 3

Frequency	1	2	3	4	5	6	7
960	1.08	0.84	0.97	0.95	0.87	1.10	0.95
900	1.21	1.40	1.29	1.92	1.74	1.60	2.46
800	1.71	2.15	1.99	7.90	8.30	7.39	0.30
720	0.69	0.79	0.75	0.67	0.89	0.99	3.36
640	0.60	0.50	0.46	0.73	0.65	0.60	13.7
600	6.58	4.52	8.21	3.20	0.71	2.21	52.6
540	1.19	1.00	0.96	1.00	1.00	1.00	0.38
480	1.31	1.66	1.30	2.40	1.44	4.31	0.60
450	2.58	2.08	2.08	1.30	1.00	1.30	1.07
400	1.77	1.08	1.62	0.39	0.33	0.33	0.64
360	1.92	1.26	2.15	1.76	1.76	1.76	1.50
320	6.67	6.79	11.2	5.00	3.00	5.51	0.78
300	3.91	3.90	3.70	25.0	15.9	27.3	2.46
270	2.53	2.85	1.15	0.69	0.69	0.69	0.79
240	1.61	0.74	2.14	2.00	2.76	2.76	8.65
225	1.33	2.67	2.11	1.36	2.63	2.24	2.68
200	0.94	0.80	0.69	1.16	0.96	0.86	0.91
180	1.27	1.04	1.09	1.45	1.15	1.36	1.02
160	1.91	0.54	0.63	3.98	0.83	0.70	0.77
150	1.11	1.11	0.81	1.25	1.33	0.81	1.22
135	1.02	0.96	1.19	1.28	1.10	1.53	1.00
120	0.83	1.53	1.19	1.20	2.86	1.36	1.70

Parts of the data shown in Table 3 are plotted in Figs. 5 and 6. Frequencies are plotted as abscissæ and intensity ratios as ordinates. Fig. 5 is for No. 1 position, Fig 6 for No. 2 position. Graphs for the other positions are very similar to these. On each plot, the dotted line, parallel to the axis of abscissæ, and representing unity ratio of intensities, is a reference line, because for any point on this line the performance of one horn is as good as that of the other.

Discussion of Results

A study of Table 3 or of the plotted results shows that, for the most part, the ratio of intensities is greater than unity, which indicates that the horn with vanes is more efficient than the one without vanes as far as intensity is concerned. This is emphasized

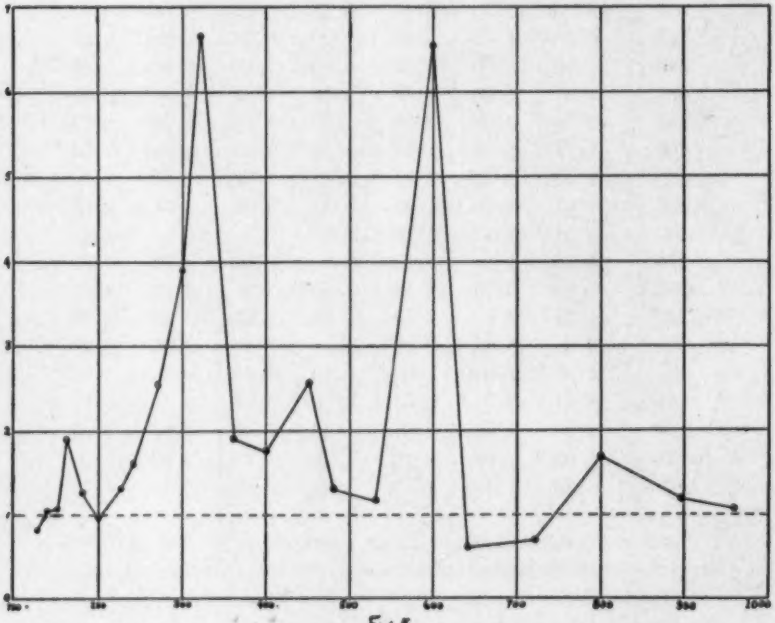


Fig 5.

in Fig. 7, in which the full graph represents the intensity from the horn with vanes, and the dotted graph that from the horn without vanes. In all but a narrow band of frequencies between 600 and 700, the horn with vanes gives the greater intensity. Data for graphs in Fig. 7 were taken for No. 4 position.

Since in most cases the intensity ratios are comparable for

different positions of the phonelescope, for the same frequency, the position of reception from either horn would seem to be immaterial.

Certain peaks on the graphs occur quite persistently about an octave apart, notably for frequencies around 300 and 600. In

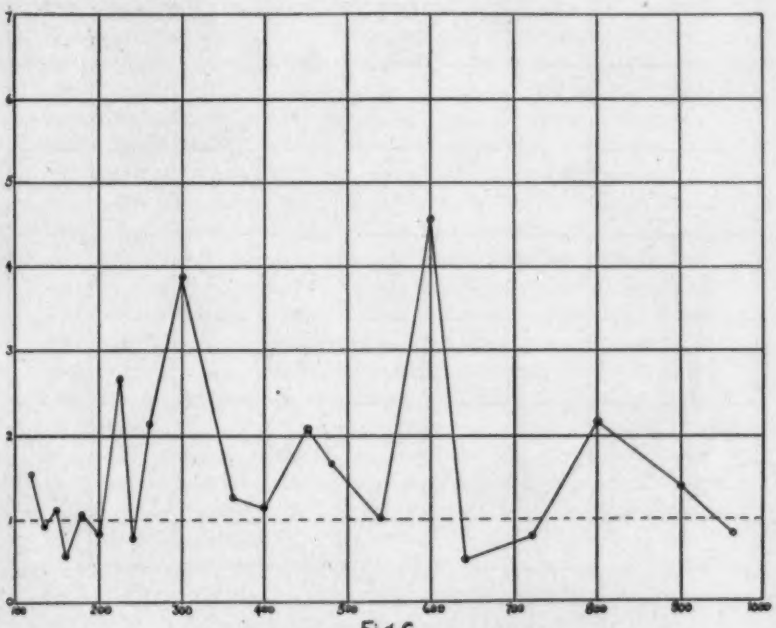


Fig 6.

some positions there is also a noticeably prominent peak at the 800 frequency. These would seem to be due to a resonance effect either in the material of the horn or in the air columns between the vanes, since the relative effects of the loud speakers and also the natural period of the receiving diaphragm have been considered. The vanes were under considerable end compression,

and were made of thin elastic material, so that certain parts of them might respond sympathetically to certain frequencies and thus cause these effects. Another possibility would be that the air columns between the vanes would act as short pipes which would resonate at the frequencies at which the peaks occur.

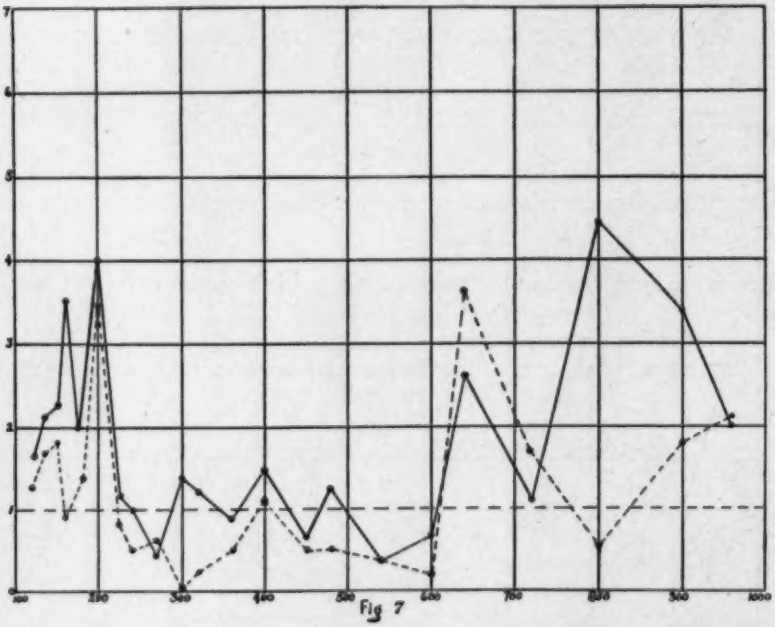


Fig 7

Conclusions

The results of the experiment indicate that for most frequencies reproduced by a phonograph, the horn with vanes is much more effective in producing volume of sound than the horn without vanes. Also, as a matter of interest only, the quality of the tone in the first case is superior to that in the second.

Rogers Laboratory of Physics,
Massachusetts Institute of Technology.

IN MEMORY OF JOSEPH LIPKA

By Norbert Wiener

The Massachusetts Institute of Technology has just suffered the loss of one of its ablest teachers and mathematicians, and this Journal of one of its most fertile contributors. Prof. Joseph Lipka died early on the morning of January 15 at the Boston Homeopathic Hospital as a consequence of pneumonia, following an operation. He was born in 1883, at Briessin in Poland. He came to the United States at an early age. In 1905 he graduated from Columbia University, where he took his master's degree in 1906. From 1907 to 1908 he served as instructor in mathematics at the University of California and in the fall of 1908 received a call to the Institute in the capacity of instructor in mathematics. In 1912, while still serving at the Institute, he completed his requirements for the doctorate at Columbia. His dissertation, entitled *Natural Families of Curves in a General Curved Space of N Dimensions* (No. 2 of the bibliography), was written under Professor Kasner, to whom he owed not merely the prevailingly geometrical trend of his mathematical interest, but more particularly his interest in the geometrical aspects of dynamical problems.

In the spring of 1913 he married Miss Jessamine McIlvane of Jamaica Plain, and spent the following summer abroad. He stayed some time at Edinburgh in contact with Whittaker. It was Whittaker's influence which turned his attention to the study of graphical and numerical methods in computation, to which Lipka's *Graphical and Mechanical Computation* (No. 6) constitutes one of the most valuable contributions in the English language. In 1914 he commenced his Mathematical Laboratory course, which was enthusiastically received at the Institute, and ultimately came to be divided into two large sections, though part of the required curriculum in only one of the Institute courses.

In 1917 Doctor Lipka was promoted to the rank of assistant professor, which he held until his promotion to an associate professorship in 1923. In 1918 he was elected a Fellow of the

American Academy of Arts and Sciences. After the publication of his *Graphical and Mechanical Computation* in 1918 began the period of his life most fruitful in pure mathematical research. From 1920 to his death there appeared from his pen no less than fourteen papers of a geometrical character. This activity received a considerable stimulus from his year in Italy and Switzerland on leave of absence in 1921–22, during which he came into close touch with many foreign scholars, especially Castelnuovo, Levi-Civita, Severi, Bompiani, and Weyl. He was the representative of the Institute at the celebration of the seven hundredth anniversary of the foundation of the University of Padua, which conferred on him the degree of the University. The influence of Levi-Civita is clearly apparent in Lipka's later writing.

Professor Lipka was an eminently successful teacher. His lecture-course on the slide-rule was a landmark at the Institute. He was active in the giving of advanced courses, not merely on geometry, but also on analytical mechanics, which he developed into one of the chief advanced courses. He combined enthusiasm, geniality, and the utmost clarity in presentation.

He had the sincere affection both of his students and of his colleagues. Even before his last voyage to Europe his health had been frail, but he went cheerfully and uncomplainingly about his daily duties, abating not one jot of his scientific activity. Death found him in the midst of a career that had yielded much and that promised much more. We have lost a valued colleague and a friend.

BIBLIOGRAPHY

1. On the Shortest Distance Between Two Consecutive Straight Lines. *Bull. Am. Math. Soc.*, V. 13 (1907), pp. 489–497.

2. Natural Families of Curves in a General Curved Space of N Dimensions. *Trans. Am. Math. Soc.*, V. 13 (1912), pp. 77–95.

3. Geometric Characterization of Isogonal Trajectories on a Surface. *Annals of Mathematics*, V. 15 (1914), pp. 145–156.

4. A Manual of Mathematics. New York, John Wiley and Sons, Inc., 1917, pp. 130.

5. Natural and Isogonal Families of Curves on a Surface. *Proc. Nat. Acad. Sci.*, V. 3 (1917), pp. 78–83.

6. Graphical and Mechanical Computation. New York, John Wiley and Sons, Inc., 1918, pp. lx+264.

7. Some Geometric Investigations on the General Problem of Dynamics. *Proc. Amer. Acad. Arts and Sciences*, V. 55 (1920), No. 7, pp. 283–322.

8. Note on Velocity Systems in Curved Space of N Dimensions. *Bull. Am. Math. Soc.*, V. 27 (1920), pp. 71–77.

9. Motion on a Surface for Any Positional Field of Force. *Proc. Nat. Acad. Sci.*, V. 6 (1920), pp. 621–624.

10. Motion on a Surface in Any Positional Field of Force. *Proc. Amer. Acad. Arts and Sciences*, V. 56 (1921), No. 4, pp. 155–181.

11. On the Geometry of Motion in Curved N-Space. *Jour. Math. Phys. M. I. T.*, V. 1 (1922), pp. 21–41.

12. Transformations of Trajectories on a Surface. *Annals of Mathematics*, V. 23 (1921), pp. 101–111.

13. Alignment Charts. *Mathematics Teacher*, April, 1921.

14. On Hamilton's Canonical Equations and Infinitesimal Contact Transformations. *Jour. Math. Phys. M. I. T.*, V. 2 (1923), pp. 31–46.

15. Sui sistemi "E" nel calcolo differenziale assoluto. *Rendiconti della reale accademia nazionale dei Lincei*, V. 31, Series 5a (1922), pp. 242–245.

16. Sulla curvatura geodetica delle linee appartinenti ad una varietà qualunque. *Lincei*, V. 31, Series 5a (1922), pp. 353–356.

17. On Irreversible Dynamical Systems. *Jour. Math. Phys. M. I. T.*, V. 2 (1923), pp. 73–87.

18. On Conformal Parallelism. *Jour. Math. Phys. M. I. T.*, V. 2 (1923), pp. 175–194.

19. Trajectory Surfaces and a Generalization of the Principal Directions in Any Space. *Proc. Amer. Acad. Arts and Sciences*, V. 59 (1923), No. 3.

20. On the Relative Curvature of Two Curves in V_n. *Bull. Am. Math. Soc.*, V. 29 (1923), pp. 345–348.

21. On Ricci's Coefficients of Rotation. *Jour. Math. Phys. M. I. T.*, V. 3 (1924), pp. 7–23.

22. (With Stanley R. Cummings.) Nomographic Charts. New York, John Wiley and Sons, Inc. In Press.

In addition numerous reviews, etc.

THE COINCIDENT POINTS OF TWO ALGEBRAIC TRANSFORMATIONS

By Frank Lauren Hitchcock

1. Let $x_1, x_2, \ldots x_n$ be a set of n independent variables. For convenience we may say that these variables are homogeneous coördinates of a point in space of $n-1$ dimensions.

Let $P_1, P_2, \ldots P_n$ be homogeneous polynomials of degree p in the variables x. We may say that these polynomials are the coördinates of a transformed point in the same space.

Similarly let $Q_1, Q_2, \ldots Q_n$ be homogeneous polynomials of degree q in the variables x. They may be said to define a second point-transformation in the space.

There exist points x such that the transformed points P and Q coincide. That is, for certain sets of values of the independent variables the corresponding values of the polynomials P are proportional to the values of the polynomials Q. This is the same as saying that the two-row determinants of the matrix

$$\begin{matrix} P_1, & P_2, & P_3, \ldots, & P_n \\ Q_1, & Q_2, & Q_3, \ldots, & Q_n \end{matrix}$$

vanish simultaneously for certain properly chosen points x.

To illustrate, if $n=3$ and $p=q=1$ we have a pair of linear transformations in a plane. There are, in general, three points altered in the same way by the two transformations, as is well known.

Again, if $n=3$, $p=2$, $q=1$, we have a linear and a quadratic transformation in a plane. There are, in general, seven points altered in the same way by the two transformations. If both transformations are quadratic with $n=3$ there are twelve such points.

I shall prove that the number of such points is in general

$$p^{n-1}+p^{n-2}q+p^{n-3}q^2+ \ldots +q^{n-1}.$$

2. The truth of the above formula has long been known for the special case $n=3$, $q=1$ when it reduces to p^2+p+1. The

proofs given by Darboux, *Bul. Sci. Math.* 13, (1878), p. 83, and by Clebsch, *Vorl. u. Geom. s.* 390, 1001, are easily extended to any q but neither proof applies to higher values of n.

I shall give a proof depending on operations called *replacements* which may be either *successive* or *simultaneous*. To make clear what is meant by a replacement, consider the case $n=4$. The six two-row determinants of the matrix

$$\begin{matrix} P_1, & P_2, & P_3, & P_4 \\ Q_1, & Q_2, & Q_3, & Q_4 \end{matrix}$$

must vanish simultaneously for the required points of space. Let us first form the three equations

$$P_1Q_2-Q_1P_2=0 \tag{1}$$

$$P_2Q_3-Q_2P_3=0 \tag{2}$$

$$P_3Q_4-Q_3P_4=0 \tag{3}$$

which are of degree $p+q$ in the independent variables x_1, x_2, x_3, x_4, and may be thought of as defining three surfaces in ordinary space. They will, in general, have $(p+q)^3$ points in common. Equations (1) and (2) are linear in P_2 and Q_2. Hence

$$P_1Q_3-Q_1P_3=0 \text{ } \textit{wherever} \text{ } P_2 \textit{ and } Q_2 \textit{ are not both zero,} \tag{4}$$

the word *wherever* applying only to the $(p+q)^3$ points. Similarly from (2) and (3)

$$P_2Q_4-Q_2P_4=0 \textit{ wherever } P_3 \textit{ and } Q_3 \textit{ are not both zero,} \tag{5}$$

and from (3) and (4)

$$P_1Q_4-Q_1P_4=0 \textit{ wherever } P_3 \textit{ and } Q_3 \textit{ are not both zero.} \tag{6}$$

Thus all six required conditions are satisfied at as many of the $(p+q)^3$ points as remain after deducting those where P_2 and Q_2 both vanish and also those where P_3 and Q_3 both vanish.

To see how many points must be deducted, note that where P_2 and Q_2 both vanish the system of equations (1), (2), (3) may be replaced by the system

$$P_2=0, \quad Q_2=0, \quad P_3Q_4-Q_3P_4=0 \tag{7}$$

satisfied at $pq(p+q)$ points. Similarly if P_3 and Q_3 both vanish the system (1), (2), (3) may be replaced by

$$P_1Q_3-Q_1P_3=0, \quad P_3=0, \quad Q_3=0 \tag{8}$$

also satisfied at $pq(p+q)$ points. Since P_2, Q_2, P_3, and Q_3 will not in general vanish together, the points (8) are all different from the points (7). In other words the two replacements are successive, not simultaneous. Therefore the six required conditions are satisfied at $(p+q)^3-2pq(p+q)$ points, which proves the formula for this case.

3. In the foregoing proof for the case $n=4$ we had to consider in succession the two systems of equations which could be obtained from the system (1) – (3) by replacing a pair of consecutive equations. In each case the original pair was replaced by a simpler pair. The new pair was obtained by equating to zero a "P" and a "Q" with the same subscript. Substituting one set of equations for another in this manner will be called a *replacement.* Thus the case $n=4$ permits two successive replacements but not two simultaneous replacements. In general the number of successive replacements will be the number of ways two consecutive objects can be chosen from $n-1$ objects which occur in a fixed order, namely $n-2$.

To see how simultaneous replacements occur, consider the case $n=5$. We begin by writing the four equations

$$P_1Q_2-Q_1P_2=0,\quad P_2Q_3-Q_2P_3=0,\quad P_3Q_4-Q_3P_4=0,\quad P_4Q_5-Q_4P_5=0$$

which are satisfied in general at $(p+q)^4$ points. By reasoning as before we see that all two-row determinants of the matrix

$$\begin{matrix} P_1, & P_2, & P_3, & P_4, & P_5 \\ Q_1, & Q_2, & Q_3, & Q_4, & Q_5 \end{matrix}$$

vanish at these points, with exception of points where one or more replacements may occur, that is, where a "P" and a "Q" having the same subscript vanish together. Thus if the first two equations are replaced by $P_2=Q_2=0$ the new system holds at $pq(p+q)^2$ points which are to be deducted from the original $(p+q)^4$ points. Similarly the replacements corresponding to $P_3=Q_3=0$ and to $P_4=Q_4=0$ occur respectively at a set of $pq(p+q)^2$ points each. If the points at which these three replacements occur were all distinct there would remain $(p+q)^4-3pq(p+q)^2$ points as the solution of the problem for the case $n=5$.

But this number is too small because the replacements $P_2=Q_2=0$ and $P_4=Q_4=0$ occur simultaneously at p^2q^2 points which have therefore been subtracted twice. The true number of points is therefore $(p+q)^4-3pq(p+q)^2+p^2q^2$ proving the formula for this case.

4. We have just seen that in the case $n=5$ two simultaneous replacements may occur. In other words the system $P_2=Q_2=P_4=Q_4=0$ is possible. For higher values of n more than two simultaneous replacements will occur.

The number of times k simultaneous replacements may occur will be the number of ways k pairs of consecutive objects can be chosen from $n-1$ objects which occur in a fixed order. It will be convenient to denote this number by $R(n-1, k)$. In particular the number of successive replacements is $R(n-1, 1)$ or $n-2$ while $R(4, 2)=1$ as in the case considered in Article 3

We are now prepared to consider the general case. I shall first show that the required number of points is given by the formula

$$(p+q)^{n-1}-R(n-1, 1)pq(p+q)^{n-3}+R(n-1, 2)p^2q^2(p+q)^{n-5}-\ \ldots$$
$$+(-1)^kR(n-1, k)p^kq^k(p+q)^{n-2k-1}+\ \ldots\ \text{etc.} \qquad \text{(A)}$$

terminating when $k=\frac{1}{2}(n-1)$ or $\frac{1}{2}(n-2)$ according as n is odd or even.

To prove this formula we note first that the term $(p+q)^{n-1}$ gives the number of points satisfying the equations

$$P_1Q_1-Q_1P_2=0,\quad P_2Q_3-Q_2P_3=0\ \ldots,\quad P_{n-1}Q_n-Q_{n-1}P_n=0$$

and that all two row determinants of the matrix of Article 1 will vanish at such of these points as permit no replacements.

Consider next those points where precisely k simultaneous replacements occur. The second term of (A) deducts all points where replacements occur, counting each replacement once. Points where k replacements may occur are therefore deducted k times by this term. The third term of (A) adds all points where two replacements occur, therefore adding points where k replacements occur as many times as two objects can be chosen from k objects or $C(k, 2)$. The fourth term of (A) deducts points where

three replacements occur, therefore points where k replacements occur will be deducted by this term $C(k, 3)$ times, and so on, up to the $(k+1)^{st}$ term of (A) which counts $(-1)^k$ times the points where k replacements occur. The remaining terms of (A) do not count these points at all. Thus the terms where precisely k replacements are possible are counted in all

$$-C(k, 1)+C(k, 2)- \ . \ . \ . \ +(-1)^k$$

times. But by the binomial theorem we have

$$(1-1)^k=0=1-C(k, 1)+C(k, 2)- \ . \ . \ . \ +(-1)^k$$

whence the points where just k replacements occur are on the whole deducted once by formula (A). The formula is therefore correct.

5. It remains to show that (A) is the same as the formula of Article 1. This follows first from the fact that

$$R(m, n)=C(m-n, n) \qquad \text{(B)}$$

To prove this, let any choice of n pairs of consecutive objects be made from m ordered objects. After the choice has been made let each pair be counted as a single object. We now have $m-n$ objects, of which n are the chosen pairs. Thus to every choice of n pairs of consecutive objects from m ordered objects there corresponds a choice of n objects from $m-n$ objects . Conversely let any selection of n objects be made from $m-n$ objects and then let every one of the n selected objects be replaced by a pair of objects. We now have a choice of n pairs of consecutive objects from m objects. The two sets of choices are thus in one to one correspondence and are therefore equal in number. Formula (A) may now be written.

$$(p+q)^{n-1}-C(n-2,1)pq(p+q)^{n-3}+C(n-3,2)p^2q^2(p+q)^{n-5},\ldots+\text{etc.}$$

which is to be shown equal to the formula of Article 1.

By proceeding as in Hall and Knight's *Higher Algebra*, page 335, we see that the expression is the coefficient of x^{n-1} in the expansion of the fraction $\dfrac{1}{(1-px)(1-qx)}$ whence the required formula

follows easily. Here the quantities a and b of Hall and Knight are to be replaced by pq and $p+q$ respectively.[1]

6. It may be remarked in conclusion that the above proof depends upon the assumption that $n-1$ equations in homogeneous coördinates are satisfied at a number of points equal to the product of their degrees. Cases of "degeneracy" and of multiple intersection may of course occur. If there are more coincident points than the formula requires there are an infinite number. Multiple points are limiting cases of the "general" case. I have elsewhere discussed in detail the case $n=3$, $p=2$, $q=1$. (A Classification of Quadratic Vectors, *Proc. Amer. Acad. of Arts and Sci.*, Vol. 52, p. 7, Jan. 1917.)

[1] The proof consists in expanding the fraction in two ways

$$\frac{1}{(1-px)(1-qx)} = \Sigma(-1)^{n-1}\frac{a^{n-1}x^{2n}}{(1-bx)^n}$$
$$= (1+px+p^2x^2+\ldots)(1+qx+q^2x^2+\ldots)$$

where n runs from 1 to ∞. Coefficients of x^{n-1} are equated.

THE QUADRATIC VARIATION OF A FUNCTION AND ITS FOURIER COEFFICIENTS

By NORBERT WIENER

At the end of an article[1] in which he gives the first example of a function of period 2π, everywhere continuous, and of limited total variation, with Fourier coefficients a_n and b_n which do not fulfil at once both conditions

$$a_n = o(1/n), \qquad b_n = o(1/n), \tag{1}$$

F. Riesz says, "In conclusion I should like to propound the following further questions which seem to me more difficult of attack: Is there a continuous function of period 2π and of limited total variation for which the sequences $\{na_n\}$, $\{nb_n\}$ converge, and at least one of the two limits is distinct from 0? Is there a discontinuous function of limited total variation for which $a_n = o(1/n)$, $b_n = o(1/n)$? The two questions are closely connected."

These questions are discussed in a paper by L. Neder,[2] who points out, in accordance with a remark of Landau, that the second question is trivial unless we restrict our question to functions with other than removable discontinuities. Under this restriction, he answers both of Riesz's questions in the negative.

Neder's results may be put together into the following theorem: the necessary and sufficient condition that a function of limited total variation and of period 2π, for which $\lim na_n$ and $\lim nb_n$ exist, should be everywhere continuous, except for removable discontinuities, is that

$$\lim na_n = 0, \quad \lim nb_n = 0. \tag{2}$$

In this paper a larger question will be discussed, the answer to which completely includes Neder's theorem. We shall show

[1] Ueber die Fourierkoeffizienten einer stetigen Function von beschränkter Schwankung, *Math. Zeitschr.* 2 (1918), p. 312.

[2] Ueber die Fourierkoeffizienten der Funktionen beschränkter Schwankung, *Math. Zeitschr.* 6 (1920), p. 270.

that if $f(x)$ is a function of limited variation with period 2π and is represented by the Fourier series

$$a_0/2+\sum_1^\infty(a_n \cos nx+b_n \sin nx),$$

which has for its complete set of discontinuities over the interval $0\leq x<2\pi$ the numbers $\{\xi_k\}$, then

$$\lim_{n\to\infty} 2n\sum_1^\infty(a_m^2+b_m^2) \sin^2 m\pi/n$$

exists and

$$=\Sigma[f(\xi_k+0)-f(\xi_k-0)]^2. \tag{3}$$

From this it may be concluded at once that a necessary and sufficient condition that $f(x)$ should be everywhere continuous, apart from removable discontinuities, is that

$$\lim_{n\to\infty} 2n\sum_1^\infty(a_m^2+b_m^2) \sin^2 m\pi/n=0. \tag{4}$$

Since we shall show that if $\lim m^2(a_m^2+b_m^2)$ exists,

$$\pi^2 \lim_{m\to\infty} m^2(a_m^2+b_m^2)=\lim_{n\to\infty} 2n\sum_1^\infty(a_m^2+b_m^2) \sin^2 m\pi/n, \tag{5}$$

Neder's theorem may be deduced at once.

In the course of this demonstration we shall introduce the notion of the quadratic variation of a function, a precise analogue to the total variation. We shall orient this and related notions with reference to one another and to Lipschitz conditions of various orders. We shall show that if a function is of limited quadratic variation over any interval, its Fourier series converges everywhere within that interval, and is uniformly convergent over any interior interval if the function is continuous over the first interval. We shall finally parallel these results in the theory of trigonometric interpolation.

§1. **The Higher Variations of a Function.** Let $f(x)$ be a single-valued, real function of the real variable x, and let it have the period 2π. Then if we write $V_s(f)$ for the upper bound of

$$|f(x_2)-f(x_1)|+|f(x_3)-f(x_2)|+\ \ldots\ +|f(x_n)-f(x_{n-1})|+|f(x_1)-f(x_n)|$$

for all sets of x_k's such that

$$0 \leq x_1 < x_2 < \ . \ . \ . \ < x_n \leq 2\pi,$$

while

$$x_k - x_{k-1} < \epsilon, \qquad (k = 2, 3, \ . \ . \ . \ , n)$$
$$x_1 + 2\pi - x_n < \epsilon, \tag{6}$$

we may define the total variation of $f(x)$ over the interval $(0, 2\pi)$ as

$$V(f) = \lim_{\epsilon \to 0} V_\epsilon(f).$$

It is understood here and in all analogous cases later that ∞ will be regarded as a possible limit or upper bound, so that $V_\epsilon(f)$ and $V(f)$ admit it as a value. There is no difficulty in showing the agreement between our definition of the total variation of a function and that more usual.

In the same way, we shall write $V_\epsilon^{(\nu)}(f)$ for the upper bound of

$$|f(x_2) - f(x_1)|^\nu + |f(x_3) - f(x_2)|^\nu + \ . \ . \ .$$
$$+ |f(x_n) - f(x_{n-1})|^\nu + |f(x_1) - f(x_n)|^\nu,$$

where the x_k's are subject to the same restrictions as in the case of $V_\epsilon(f)$. We shall then define the νth total variation of $f(x)$ as

$$V^{(\nu)}(f) = \lim_{\epsilon \to 0} V_\epsilon^{(\nu)}(f). \tag{7}$$

It is easy to show that if O_m is the oscillation of $f(x)$ in the interval (x_m, x_{m+1}), and O_n is the oscillation of $f(x)$ in the interval $(x_n, x_1 + 2\pi)$, then

$$V^{(\nu)}(f) = \lim_{\epsilon \to 0} \lim \sup \sum_1^n O_m^\nu, \tag{8}$$

where the lim sup is taken over all the sets of O_m's corresponding to sequences of x_k's satisfying (6). It is obvious on inspection that

$$\lim_{\epsilon \to 0} \lim \sup \sum_1^n O_m^\nu \geq V^{(\nu)}(f), \tag{9}$$

since the oscillation of a function in an interval is at least as great as the difference of the values of the function at the ends of the interval. On the other hand, ΣO_m^ν is the upper bound of

$$|f(\xi_1) - f(\eta_1)|^\nu + |f(\xi_2) - f(\eta_2)|^\nu +$$
$$+ |f(\xi_{n-1}) - f(\eta_{n-1})|^\nu + |f(\xi_n) - f(\eta_n)|^\nu,$$

where

$$x_k \leq \xi_k \leq x_{k+1}, \qquad x_k \leq \eta_k \leq x_{k+1}, \qquad (k=1, 2, \ldots, n-1)$$
$$x_n \leq \xi_n \leq 2\pi + x_1, \qquad x_n \leq \eta_n \leq 2\pi + x_1.$$

It results that

$$\Sigma O_m^{\nu} \leq V_\epsilon^{(\nu)}(f), \tag{10}$$

which in combination with (9) yields (8) almost at once.

If $V^{(\nu)}(f)$ is bounded, $f(x)$ can have no discontinuities not of the first kind. Otherwise we shall have some point ξ at which either

$$\bar{f}(\xi+0) > \underline{f}(\xi+0),$$

or

$$\bar{f}(\xi-0) > \underline{f}(\xi-0),$$

or

$$\bar{f}(\xi \pm 0) = \infty,$$

or

$$\underline{f}(\xi \pm 0) = -\infty.$$

In the last two cases, it is possible, given any A, to find in any neighborhood of ξ numbers ξ_1 and ξ_2 such that $f(\xi_2) - f(\xi_1) > A$. Hence $V^{(\nu)}(f) \geq A^{\nu}$, and is therefore infinite. Excluding these, in the first case in any neighborhood to the right of ξ there is a number ξ_1 such that

$$|\bar{f}(\xi+0) - f(\xi_1)| < \eta.$$

Between ξ and ξ_1 there is a number ξ_2 such that

$$|\underline{f}(\xi+0) - f(\xi_2)| < \eta.$$

Between ξ and ξ_2 there is a number ξ_3 such that

$$|\bar{f}(\xi+0) - f(\xi_3)| < \eta,$$

and so on indefinitely. Let us suppose that $2\eta < \bar{f}(\xi+0) - \underline{f}(\xi+0)$, and that $\xi_1 - \xi < \epsilon$. Then

$$V_\epsilon^{(\nu)}(f) \geq \sum_1^n |f(\xi_{n+1}) - f(\xi_n)|^{\nu}$$
$$\geq n\{\bar{f}(\xi+0) - \underline{f}(\xi+0) - 2\eta\}^{\nu}. \tag{11}$$

Since n may have any value whatever, independently of η and ϵ, we find that $V_\epsilon^{(\nu)}(f)$ is infinite for all ϵ, so that $V^{(\nu)}(f)$ is infinite.

An argument along identical lines excludes the possibility that $f(\xi-0)$ does not have a definite finite value. Thus $f(x)$ has only discontinuities of the first kind, which by a familiar theorem must form a denumerable set. Here and throughout this paper we shall call these discontinuities $\xi_1, \xi_2, \ldots, \xi_k, \ldots$, modulo 2π. For the sake of simplicity of notation we shall confine ourselves throughout the remainder of the paper to functions such that for every k, $f(\xi_k)$ lies between $f(\xi_k-0)$ and $f(\xi_k+0)$. The modifications to be made in our argument in the contrary case will be obvious.

Let ϵ be any number less than the least distance (modulo 2π) between any tw oof the points ξ_k ($k \leq m$), and such that if $\epsilon_1 \leq \epsilon$, then for all $k \leq m$,

$$|f(\xi_k-\epsilon_1)-f(\xi_k-0)|<\eta;$$
$$|f(\xi_k+\epsilon_1)-f(\xi_k+0)|<\eta.$$

Then if $x_1, \ldots, x_n=x_0$ is any set of numbers such that the residue modulo 2π of x_i-x_{i-1} lies between 0 and ϵ for all i, we have

$$\sum_1^n |f(x_i)-f(x_{i-1})|^\nu \geq \sum_1^m |f(\xi_k+0)-f(\xi_k-0)-2\eta|^\nu.$$

Since an ϵ can be found for any η, no matter how small, and since if ϵ is associated with η, we may associate with η any smaller number ϵ', it follows that no matter how the x_i's are chosen for each value of ϵ,

$$\lim_{\epsilon \to 0} \sum_1^n |f(x_i)-f(x_{i-1})|^\nu \geq \sum_1^m |f(\xi_k+0)-f(\xi_k-0)|^\nu.$$

Since this is true for all m, we have

$$\lim_{\epsilon \to 0} \sum_1^n |f(x_i)-f(x_{i-1})|^\nu \geq \sum_1^\infty |f(\xi_k+0)-f(\xi_k-0)|^\nu. \qquad (12)$$

§2. **Relations between Variations of the νth and of Lower Orders.** By Taylor's Theorem with Remainder, if ν is a positive number greater than or equal to 2,

$$v^\nu = u^\nu + \nu(v-u)\{u+\theta(v-u)\}^{\nu-1},$$

where θ is some number between 0 and 1, inclusive. It follows that if u and v are both positive,

$$|v^\nu - u^\nu| \leq \nu(v-u)(v^{\nu-1}+u^{\nu-1}).$$

From this we may deduce at once that if H is the maximum fluctuation made by $f(x)-g(x)$ in an interval of length ϵ, then

$$|V_\epsilon^{(\nu)}(f) - V_\epsilon^{(\nu)}(g)| \leq \nu H(V_\epsilon^{(\nu-1)}(f) + V_\epsilon^{(\nu-1)}(g)). \tag{13}$$

Let $f(x)$ be a function of limited $(\nu-1)$st variation V, with discontinuities $\{\xi_k\}$. Let the saltus of f at ξ_k be a non-increasing function of k. Let $|f(\xi_n+0)-f(\xi_n-0)| < \eta$. Let $g(\xi_\nu+0) = f(\xi_\nu+0)$ and $g(\xi_\nu-0) = f(\xi_\nu-0)$ $[\nu = 1, 2, \ldots, n]$, and let $g(x)$ be defined elsewhere by linear interpolation between these values. Since $V^{(\nu-1)}(g) \leq V^{(\nu-1)}(f)$, it follows that for sufficiently small values of ϵ,

$$|V_\epsilon^{(\nu)}(f) - V_\epsilon^{(\nu)}(g)| \leq 4\nu\eta V. \tag{14}$$

That is,

$$|V_\epsilon^{(\nu)}(f) - \sum_1^n |f(\xi_k+0)-f(\xi_k-0)|^\nu| \leq 4\nu\eta V.$$

From this it may be deduced at once that

$$\lim_{n\to\infty} |V^{(\nu)}(f) - \sum_1^\infty |f(\xi_k+0)-f(\xi_k-0)|^\nu|.$$

$$\leq \lim_{n\to\infty} \{\sum_{n+1}^\infty |f(\xi_k+0)-f(\xi_k-0)|^\nu + 4\nu\eta V\}$$

$$\leq \lim_{\eta\to 0} \eta V(4\nu+1) = 0. \tag{15}$$

Hence

$$V^{(\nu)}(f) = \sum_1^\infty |f(\xi_k+0)-f(\xi_k-0)|^\nu. \tag{16}$$

An immediate consequence of this is that a continuous function of bounded $(\nu-1)$st variation is of νth variation 0. However, this results admits of considerable extension, since if $\nu_1 < \nu$,

$$V^{(\nu)}(f) \leq (\limsup |f(x+0)-f(x-0)|)^{\nu-\nu_1} V^{(\nu_1)}(f). \tag{17}$$

Thus if $f(x)$ is of period 2π and everywhere continuous, and is of limited ν_1st variation, it is of νth variation 0 for $\nu > \nu_1$.

§3. **Special Theorems Relating to the Quadratic Variation.** Clearly

$$V^{(2)}_{\frac{2\pi}{n}}(f) \geq \left[f\left(u+\frac{2\pi}{n}\right)-f(u)\right]^2+\left[f\left(u+\frac{4\pi}{n}\right)-f\left(u+\frac{2\pi}{n}\right)\right]^2+\cdots + \left[f(u+2\pi)-f\left(u+2\pi\left(\frac{n-1}{n}\right)\right)\right]^2. \tag{18}$$

Now let $f(x)$ be summable and of summable square between 0 and 2π. Then there will be a Fourier series

$$a_0/2+\sum_1^\infty (a_n \cos nx+b_n \sin nx)$$

which converges in the mean to $f(x)$ over $(0, 2\pi)$. Then

$$f\left(u+\frac{2\pi}{n}\right)-f(u)\sim\sum_1^\infty \left\{a_m\left[\cos m\left(u+\frac{2\pi}{n}\right)-\cos mu\right] + b_m\left[\sin m\left(u+\frac{2\pi}{n}\right)-\sin mu\right]\right\}$$

$$\sim\sum_1^\infty\left[2b_m \cos m\left(u+\frac{\pi}{n}\right)-2a_m \sin m\left(u+\frac{\pi}{n}\right)\right]\sin\frac{m\pi}{n}. \tag{19}$$

Hence

$$\int_0^{2\pi}\left[f\left(u+\frac{2\pi}{n}\right)-f(u)\right]^2 du = 4\pi\sum_1^\infty (a_m^2+b_m^2)\sin^2\frac{m\pi}{n}. \tag{20}$$

Therefore by (18),

$$V^{(2)}_{\frac{2\pi}{n}}(f) \geq \frac{n}{2\pi}\int_0^{\frac{2\pi}{n}}\left\{\left[f\left(u+\frac{2\pi}{n}\right)-f(u)\right]^2+\cdots + \left[f(u+2\pi)-f\left(u+2\pi\left(\frac{n-1}{n}\right)\right)\right]^2\right\}du$$

$$=\frac{n}{2\pi}\int_0^{2\pi}\left[f\left(u+\frac{2\pi}{n}\right)-f(u)\right]^2 du$$

$$=2n\sum_1^\infty (a_m^2+b_m^2)\sin^2\frac{m\pi}{n}. \tag{21}$$

Combining this with (16), we see that if $f(x)$ is of limited total variation,

$$\overline{\lim_{n\to\infty}}\ 2n\sum_1^\infty (a_m^2+b_m^2)\sin^2\frac{m\pi}{n} \leq \sum_1^\infty [f(\xi_k+0)-f(\xi_k-0)]^2. \tag{22}$$

The expression on the left hand of this inequality we shall term $Q(f)$. Moreover, since an integral lies between the extremes of the integrand, there is a u such that

$$\left[f\left(u+\frac{2\pi}{n}\right)-f(u)\right]^2+\ \ldots\ +\left[f(u+2\pi)-f\left(u+2\pi\left(\frac{n-1}{n}\right)\right)\right]^2 \leq \frac{n}{2\pi}\int_0^{2\pi}\left[f\left(u+\frac{2\pi}{n}\right)-f(u)\right]^2 du. \tag{23}$$

It follows from (12), (21), and (23) that

$$\lim_{n\to\infty} 2n\sum_1^\infty(a_m^2+b_m^2)\sin^2\frac{m\pi}{n} \geq \sum_1^\infty[f(\xi_k+0)-f(\xi_k-0)]^2. \tag{24}$$

(22) and (24) together give:

$$\lim_{n\to\infty} 2n\sum_1^\infty(a_m^2+b_m^2)\sin^2\frac{m\pi}{n} = \sum_1^\infty[f(\xi_k+0)-f(\xi_k-0)]^2. \tag{3}$$

If $f(x)$ is everyhere continuous except for removable singularities, which are nugatory in so far as the Fourier coefficients are concerned, (4) follows at once.

The author has already shown in a paper forthcoming in the *American Journal of Mathematics* that if s_m is the mth partial sum of a series, the limit

$$\lim_{n\to\infty}\frac{2n}{\pi^2}\sum_1^\infty\frac{s_m}{m^2}\sin^2\frac{m\pi}{n}$$

may be regarded as a generalized sum of this series which is linear and regular in the sense of W. A. Hurwitz.[3] We shall develop some of the results of that paper in order to prove (5).

Let us assume that for all m

$$a_m^2+b_m^2 \geq A/m^2.$$

Then for all n,

$$\sum_1^\infty[f(\xi_k+0)-f(\xi_k-0)]^2 \geq 2An\sum_1^\infty\frac{1}{m^2}\sin^2\frac{m\pi}{n} = An\sum_1^\infty\left(\cos 0-\cos\frac{2m\pi}{n}\right)\frac{}{m^2} = A\pi^2(1-1/n). \tag{25}$$

[3] Bull. Am. Math. Soc., Vol. 28 (1922), p. 18.

A similar result holds with the sign of inequality reversed. Moreover, no batch of terms at the beginning of the a_m's and b_m's is of any effect in determining $\lim_{n\to\infty} 2n\sum_1^\infty (a_m^2+b_m^2)\sin^2 m\pi/n$. Hence

$$\varlimsup_{m\to\infty} \pi^2 m^2(a_m^2+b_m^2) \leq \sum_1^\infty [f(\xi_k+0)-f(\xi_k-0)]^2 \leq \varlimsup_{m\to\infty} \pi^2 m^2(a_m^2+b_m^2). \tag{26}$$

(5) follows at once.

Again, if $0 < x \leq \pi/2$, $\sin x \geq 2x/\pi$. Therefore

$$\begin{aligned}\sum_1^\infty [f(\xi_k+0)-f(\xi_k-0)]^2 &\geq 4n\sum_1^\infty (a_m^2+b_m^2)\sin^2 m\pi/2n \\ &\geq 4n\sum_1^n (a_m^2+b_m^2)\sin^2 m\pi/2n \\ &\geq 4n\sum_1^n (a_m^2+b_m^2)m^2/n^2 \\ &= \frac{4}{n}\sum_1^n m^2(a_m^2+b_m^2). \end{aligned}\tag{27}$$

Therefore, if $f(x)$ is a continuous function of limited total variation and period 2π, the Cesàro limit of first order of $m^2(a_m^2+b_m^2)$ is 0.

Now let $f(x)$ be a function of limited total variation and period 2π, and let

$$\lim_{n\to\infty} \sum_1^n \frac{m^2}{n}(a_m^2+b_m^2) = 0.$$

By a well-known theorem concerning functions of limited total variation,

$$a_n = 0(1/n), \qquad b_n = 0(1/n).^4$$

Let the upper bound of $m^2(a_m^2+b_m^2)$ be A. Then

$$\begin{aligned}2n\sum_1^\infty (a_m^2+b_m^2)\sin^2\frac{m\pi}{n} &\leq 2n\sum_1^{np} (a_m^2+b_m^2)\frac{m^2\pi^2}{n^2} + 2n\sum_{np+1}^\infty (a_m^2+b_m^2) \\ &\leq \frac{2\pi^2}{n}\sum_1^{np} m^2(a_m^2+b_m^2) + 2nA\sum_{np+1}^\infty \frac{1}{m^2} \\ &\leq \frac{2\pi^2}{n}\sum_1^{np} m^2(a_m^2+b_m^2) + \frac{2A}{p}. \end{aligned}\tag{28}$$

Cf. Riesz, loc. cit.

If we let n become infinite, we get:

$$\lim_{n\to\infty} 2n\sum_{1}^{\infty}(a_m^2+b_m^2)\sin^2\frac{m\pi}{n}\leq\frac{2A}{p}. \tag{29}$$

Since p is arbitrary, (3) yields us immediately:

$$\sum_{1}^{\infty}[f(\xi_k+0)-f(\xi_k-0)]^2=0. \tag{30}$$

Joining this result with that of the last paragraph, we obtain the highly important theorem: *if $f(x)$ is of limited total variation and of period 2π, the necessary and sufficient condition that it be continuous is that*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{1}^{n}m^2(a_m^2+b_m^2)=0. \tag{31}$$

§4. **Higher Variations and Lipschitz Conditions.** We shall say that $f(x)$ satisfies a Lipschitz condition of order $1/\nu$ at $x=\xi$ if whenever ϵ is a number less in modulus than some positive E, there is an A such that

$$|f(\xi+\epsilon)-f(\xi)|\leq A\epsilon^{1/\nu}. \tag{32}$$

on the other hand, we shall say that $f(x)$ satisfies a Lipschitz condition of order $1/\nu$ uniformly over (a, b) if there is an A such that for all x and y on the closed interval (a, b)

$$|f(x)-f(y)|\leq A|x-y|^{1/\nu}. \tag{33}$$

It is clear that if $f(x)$ is a function of period 2π satisfying (33) over an interval containing in its interior $(0, 2\pi)$, then

$$V^{(\nu)}(f)\leq 2\pi A.$$

Again, let $f(x)$ be such that $V^{(\nu_1)}(f)$ is finite for some $\nu_1<\nu$. Then some $V_\epsilon^{(\nu_1)}(f)$, and hence every $V_\epsilon^{(\nu_1)}(f)$, will be finite. Let $B=V_{2\pi}^{(\nu_1)}(f)$. Divide the interval $(0, 2\pi)$ into consecutive halves, fourths, . . . , $1/2^n$ths. Let the kth part of length $2\pi/2^n$ be I_{kn}, and let the fluctuation of f on the closed interval I_{kn} be F_{kn}. Then

$$B\geq\sum_{k=1}^{2n}F_{kn}^{\nu_1}. \tag{34}$$

From this point on we shall use the symbol $[x]$ for the largest integer not greater than x. It results from (34) that over not more than $1+[B/a_n^{\nu_1}]$ of the I_{kn}'s can the fluctuation of f exceed a_n. Let

$$a_n = C(2\pi/2^n)^{1/\nu},$$

where C and ν are independent of n. Then the total length L_n of the I_{kn}'s for which

$$F_{k-1,n} > C(2\pi/2^n)^{1/\nu},\ F_{kn} > C(2\pi/2^n)^{1/\nu},\ \text{or}\ F_{k+1,n} > C(2\pi/2^n)^{1/\nu}$$ [8]

is not more than

$$3(2\pi/2^n)\left\{[B/(C^{\nu_1}(2\pi/2^n)^{\nu_1/\nu})]+1\right\}.$$

Hence for sufficiently large values of n,

$$L_n < 6B(2\pi)^{1-\nu_1/\nu}/\left\{2^{1-\nu_1/\nu}\right\}^n C^{\nu_1}. \tag{35}$$

It follows that there is a number N such that

$$\sum_N^\infty L_n < 6B(2\pi)^{1-\nu_1/\nu} \sum_N^\infty \left\{2^{\nu_1/\nu-1}\right\}^n / C^{\nu_1}$$
$$< D/C^{\nu_1}, \tag{36}$$

where D is a constant independent of C.

As a consequence, there is a set of points S_C of measure at least $2\pi - D/C^{\nu_1}$ with the property that if ξ belongs to S_C, then for any integer n from a certain N on, ξ lies in the middle third of an interval of length $6\pi/2^n$, over which the fluctuation of f is not greater than $3C(2\pi/2^n)^{1/\nu}$. In other words, if $\epsilon < 2\pi/2^N$, we have

$$|f(\xi+\epsilon) - f(\xi)| \leq 3C(2\epsilon)^{1/\nu}$$
$$\leq 3C2^{1/\nu}\epsilon^{1/\nu}. \tag{37}$$

That is, at any point in any S_C, $f(x)$ satisfies a Lipschitz condition of order $1/\nu$. However, ΣS_C is a measurable set with a measure $\geq 2\pi - D/C^{\nu_1}$, whatever C may be. Thus if $\nu_1 > 0$, ΣS_C is of measure 2π, and its complement is a set of zero measure. Consequently, if $V^{(\nu_1)}(f)$ is bounded and $0 < \nu_1 < \nu$, $f(x)$ satisfies a Lipschitz condition of order $1/\nu$ except over a set of zero measure. From this the further important conclusion may be drawn that

[8] Here $F_{0,n} = F_{2^n,n}$.

the Fourier series of a function of limited νth variation converges except at a set of points of zero measure.

§5. **The Convergence of the Fourier Series of a Function of Limited Quadratic Variation.** Let $s_n(x)$ be the nth partial sum of the Fourier series of $f(x)$, and let $\sigma_n(x)$ be $(s_1+ \ . \ . \ . \ +s_n)/n$, the nth approximation to the first Cesàro sum of the Fourier series of f. Then by the Schwarz inequality,

$$\begin{aligned}|s_{n+1}(x)-\sigma_n(x)| &= \{(a_1\cos x+b_1\sin x)+2(a_2\cos 2x+b_2\sin 2x)\\ &\quad + \ . \ . \ . \ +n(a_n\cos nx+b_n\sin nx)\}/n\\ &\leq \sqrt{(a_1^2+b_1^2+4a_2^2+4b_2^2+ \ . \ . \ . \ +n^2a_n^2+n^2b_n^2)\times}\\ &\quad \sqrt{(\cos^2 x+\sin^2 x+ \ . \ . \ . \ +\cos^2 nx+\sin^2 nx)/n^2}\\ &= \sqrt{\sum_1^n \frac{\nu^2}{n}(a_\nu^2+b_\nu^2)}\\ &\leq \sqrt{4n\sum_1^\infty (a_\nu^2+b_\nu^2)\sin^2\frac{m\pi}{2n}}\\ &= \sqrt{\frac{n}{\pi}\int_0^{2\pi}\left\{f\left(x+\frac{\pi}{n}\right)-f(x)\right\}^2 dx}. \end{aligned} \tag{38}$$

Hence if $f(x)$ has a Fourier series summable $(C1)$ at any point (which will always be true if $V^{(2)}(f)\neq\infty$ and $f(x)$ has consequently only singularities of the first kind), and if $Q(f)=0$, $f(x)$ will converge at the point where the Fourier series is summable. Indeed, it will converge uniformly over any interval over which the Fourier series is uniformly summable, and hence over any interval contained in an interval over which $f(x)$ is continuous. Furthermore, if the Fourier series of $f(x)$ is summable at $x=x_1$ and $Q(f)$ is bounded, the nth partial sum of the Fourier series of $f(x)$ at x_1 lies in the interval

$$(f(x_1)-\sqrt{Q(f)}-\epsilon_n,\ f(x_1)+\sqrt{Q(f)}+\epsilon_n)$$

where ϵ_n tends to 0 with $1/n$, and tends to 0 uniformly over any interval over which $f(x_1)$ is uniformly summable.

It is well known that the convergence and uniform convergence

properties of a Fourier series are entirely local properties of the function it represents. In more precise language, if $f(x)$ and $g(x)$ are two summable functions coinciding over the interval (a, b), and if (c, d) is an interval interior to (a, b), then the Fourier series for $f-g$ converges uniformly to 0 over (c, d). Therefore if $f(x)$ is represented by a Fourier series summable at x_1, the nth sum of the Fourier series for $f(x)$ at x_1 lies in the interval

$$(f(x_1)-\sqrt{Q(g)}-\epsilon_n,\ f(x_1)+\sqrt{Q(g)}+\epsilon_n)$$

where ϵ_n vanishes with $1/n$, and vanishes uniformly over any interval interior to (c, d) over which the Fourier series for $f(x_1)$ is uniformly summable.

Let $f(x)$ have no discontinuities of the second kind. Let us then put:

$$g(x)=f(x); \qquad (a<x<b)$$

$$g(x)=f(b-0)+\frac{x-b}{2\pi+a-b}f(a+0); \quad (b\leq x\leq 2\pi+a)$$

$$g(x+2\pi)=g(x).$$

We shall formulate the definitions:

$$\overset{b}{\underset{a}{Q}}(f)=Q(g), \quad \overset{b}{\underset{a}{V}}{}^{(2)}_{*}(f)=V^{(2)}_{*}(g), \text{ and } \overset{b}{\underset{a}{V}}{}^{(2)}(f)=V^{(2)}(g). \tag{39}$$

From these it immediately follows that if $0<a<c<b<2\pi$,

$$\overset{b}{\underset{a}{Q}}(f)\leq\overset{c}{\underset{a}{Q}}(f)+\overset{b}{\underset{c}{Q}}(f)+\{f(c+0)-f(c-0)\}^2, \tag{40}$$

and

$$\overset{b}{\underset{a}{V}}{}^{(2)}(f)=\overset{c}{\underset{a}{V}}{}^{(2)}(f)+\overset{b}{\underset{c}{V}}{}^{(2)}(f)+\{f(c+0)-f(c-0)\}^2. \tag{41}$$

Further obvious relations are

$$\overset{b}{\underset{a}{V}}{}^{(2)}(f)\geq\overset{b}{\underset{a}{Q}}(f) \tag{42}$$

and

$$\overset{b}{\underset{a}{Q}}(f)\leq Q(f), \quad \overset{b}{\underset{a}{V}}{}^{(2)}(f)\leq V^{(2)}(f). \tag{43}$$

Among the corollaries of these statements are:

(a) $\overset{x}{\underset{a}{V}}{}^{(2)}(f)$ is an increasing function of x, and as such has

only a denumerable set of discontinuities, provided that $V^{(2)}(f)$ exists;

(*b*) If ξ is a point of continuity of $\overset{x}{\underset{a}{V}}{}^{(2)}(f)$, $\overset{b_1}{\underset{a_1}{V}}{}^{(2)}(f)$ and hence $\overset{b_1}{\underset{a_1}{Q}}(f)$ can be made arbitrarily small for some interval (a_1, b_1) containing ξ, so that the Fourier series of $f(x)$ must converge to $f(x)$ at ξ.

It remains to discuss the continuity of $\overset{x}{\underset{a}{V}}{}^{(2)}(f)$. We shall show that if $\overset{b}{\underset{a}{V}}{}^{(2)}(f)$ is finite, $\overset{x}{\underset{a}{V}}{}^{(2)}(f)$ is continuous within (a, b) and the Fourier series of $f(x)$ converges to $f(x)$ except at the discontinuities of $f(x)$. For let $\overset{x}{\underset{a}{V}}{}^{(2)}(f)$ be discontinuous at ξ within (a, b), while $f(x)$ is continuous. It follows that no matter how small ϵ is, $\overset{\xi+\epsilon}{\underset{\xi-\epsilon}{V}}{}^{(2)}(f)$ remains greater than some positive number A. That is, given any positive ϵ and η, it is possible to find a sequence of numbers $\{x_k\}$ such that

$$\xi-\epsilon\leq x_1\leq \ . \ . \ . \ \leq x_n\leq\xi+\epsilon;$$
$$|x_k-x_{k-1}|<\eta; \quad (k=2, \ . \ . \ . \ , n)$$
$$\sum_{2}^{n}\{f(x_k)-f(x_{k-1})\}^2>A.$$

Clearly ξ will lie in the interior of some interval (x_{j-1}, x_{j+1}). Since $f(x)$ is continuous at ξ, we can make $\{f(x_j)-f(x_{j-1})\}^2 + \{f(x_{j+1})-f(x_j)\}^2$ as small as we choose by taking η sufficiently small. Thus we can make

$$\sum_{2}^{j-1}\{f(x_k)-f(x_{k-1})\}^2+\sum_{j+2}^{n}\{f(x_k)-f(x_{k-1})\}^2>A.$$

Repeating this process, we can find a sequence of numbers $\{y_k\}$ such that

$$x_{j-1}\leq y_1 \ \leq \ . \ . \ . \ \leq y_m\leq x_{j+1};$$
$$y_{k-1}\leq \xi \leq y_{k+1},$$

and

$$\sum_{2}^{k-1}\{f(y_k)-f(y_{k-1})\}^2+\sum_{k+2}^{m}\{f(y_k)-f(y_{k-1})\}^2>A.$$

This means that $\overset{b_1}{\underset{a_1}{V_\epsilon^{(2)}}}(f) > 2A$, where ϵ is arbitrarily small, and hence that $\overset{b_1}{\underset{a_1}{V^{(2)}}}(f) \geq 2A$. In the same way it may be shown that for any N, $\overset{b_1}{\underset{a_1}{V^{(2)}}}(f) \geq NA$ and hence $\overset{b_1}{\underset{a_1}{V^{(2)}}}(f) = \infty$. However, this contradicts our initial assumption.

It follows that if $\overset{b_1}{\underset{a_1}{V^{(2)}}}(f)$ is bounded and (a, b) is an interval interior to (a_1, b_1), $\overset{x}{\underset{a}{V^{(2)}}}(f)$ is *uniformly* continuous over (a,b). That is, given any small quantity θ, (a, b) may be included in a finite number of overlapping intervals $(\alpha_\epsilon, \beta_\epsilon)$, over each of which

$$\overset{\beta_\epsilon}{\underset{\alpha_\epsilon}{V^{(2)}}}(f) < \theta.$$

Since there is only a finite set of these, it follows that the range of oscillation of the nth sum of the Fourier series for $f(x)$ is included in the interval

$$(f(x)-\sqrt{\theta}-\epsilon_n,\ \ f(x)+\sqrt{\theta}+\epsilon_n),$$

where ϵ_n vanishes uniformly with $1/n$ over (a, b). We may conclude directly from this the uniform convergence of the Fourier series for $f(x)$ over (a, b) to the value of the function.

If $\overset{b}{\underset{a}{V^{(2)}}}(f)$ is bounded and $f(x)$ is discontinuous at $\xi[a<\xi<b]$, the Fourier series of f converges to $\frac{1}{2}\{f(\xi+0)+f(\xi-0)\}$ at ξ. To show this, it is only necessary to remark that $f(x)$ has only a discontinuity of the first kind at ξ. Let $g(x)$ be defined as follows:

$$\begin{aligned}
&g(\xi)=0;\\
&g(x)=f(x)-f(\xi-0); \qquad a<x<\xi\\
&g(x)=f(x)-f(\xi+0); \qquad \xi<x<b\\
&g(x)=f(x) \text{ otherwise.}
\end{aligned}$$

Clearly $g(x)$ is continuous at $x=\xi$, and

$$\overset{b}{\underset{a}{V^{(2)}}}(g) \leq \overset{b}{\underset{a}{V^{(2)}}}(f)+2\{f(\xi+0)-f(\xi-0\}^2.$$

Hence the Fourier series of $g(x)$ converges to 0 for $x=\xi$. More-

over, $f(x)-g(x)$ is a function of limited total variation, and its Fourier series converges to

$$\tfrac{1}{2}\{f(\xi+0)-g(\xi+0)+f(\xi-0)-g(\xi-0)\}$$
$$=\tfrac{1}{2}\{f(\xi+0)+f(\xi-0)\}$$

for $x=\xi$. Adding these series term by term, we obtain our result.

§6. **The Relation between $Q(f)$ and $V^{(2)}(f)$.** The close relation between $Q(f)$ and $V^{(2)}(f)$ might lead us to believe that the boundedness of the former is a sufficient condition for the boundedness of the latter. That this is not the case will be shown by the following example:

$$f(x)=\sin\log x; \qquad 0<x\leq 2\pi$$
$$f(x+2\pi)=f(x).$$

Since $f(+0)$ is indeterminate, $V^{(2)}(f)$ cannot be finite. On the other hand,

$$\begin{aligned}
Q(f)&=\overline{\lim_{n\to\infty}}\frac{n}{2\pi}\left\{\int_0^{2\pi-\frac{2\pi}{n}}\left\{\sin\log\left(u+\frac{2\pi}{n}\right)-\sin\log u\right\}^2du\right.\\
&\qquad\left.+\int_{2\pi-\frac{2\pi}{n}}^{2\pi}\left\{\sin\log\left(u+\frac{2\pi}{n}-2\pi\right)-\sin\log u\right\}^2du\right\}\\
&\leq\overline{\lim_{n\to\infty}}\frac{n}{2\pi}\left\{\int_0^{\frac{2\pi}{n}}4\,du\right.\\
&\qquad+\int_{\frac{2\pi}{n}}^{2\pi-\frac{2\pi}{n}}\left\{2\cos\log\left[u\left(u+\frac{2\pi}{n}\right)\right]\sin\log\left(1+\frac{2\pi}{nu}\right)\right\}^2du\\
&\qquad\left.+\int_{2\pi-\frac{2\pi}{n}}^{2\pi}4\,du\right\}\\
&\leq 8+\overline{\lim_{n\to\infty}}\frac{2n}{\pi}\int_{\frac{2\pi}{n}}^{2\pi-\frac{2\pi}{n}}\sin\log\left(1+\frac{2\pi}{nu}\right)du\\
&\leq 8+\overline{\lim_{n\to\infty}}\frac{2}{\pi}\int_{2\pi}^{2\pi(n-1)}\left[\log\left(1+\frac{2\pi}{x}\right)\right]^2dx\\
&\leq 8+\frac{2}{\pi}\int_{2\pi}^{\infty}\frac{4\pi^2}{x^2}\,dx=12.
\end{aligned}$$

§7. **An Expression Related to $Q(f)$.** If $Q(f)$ is bounded, the expression

$$\frac{\int_0^\infty 2n \sum_{m=1}^{\infty} (a_m^2+b_m^2)\sin^2 \frac{m\pi}{n} \cdot \frac{e^{-\frac{k}{n}}}{n^2} dn}{\int_0^\infty \frac{e^{-\frac{k}{n}}}{n^2} dn}$$

exists for every value of k, and has the upper bound of its ultimate oscillation not greater than $Q(f)$. That is,

$$Q(f) \geq \overline{\lim_{k\to\infty}} \frac{\int_0^\infty 2\sum_1^\infty (a_m^2+b_m^2) \sin^2 m\pi\xi \, e^{-\xi k} d\xi}{\int_0^\infty \xi e^{-\xi k} d\xi}$$

$$= \overline{\lim_{k\to\infty}} \sum_1^\infty \frac{(a_m^2+b_m^2)m^2\pi^2 k}{k^2+m^2\pi^2}. \tag{44}$$

It follows that

$$Q(f) \geq \overline{\lim_{\mu\to\infty}} \sum_\mu^\infty \frac{(a_m^2+b_m^2)m^2\pi^2\mu}{\pi^2(m^2+\mu^2)} \geq \overline{\lim_{\mu\to\infty}} \sum_\mu^\infty \frac{\pi\mu}{2}(a_m^2+b_m^2), \tag{45}$$

this giving us a measure of the rapidity of the convergence in the mean of the Fourier series of a function with bounded Q. What is more, we can proceed directly from the series of the Poisson integral to a proof of the convergence of a Fourier series of a function of limited quadratic variation. We have

$$\left| a_0/2 + \sum_1^\infty \left(1-\frac{1}{n}\right)^m (a_m \cos mx + b_m \sin mx) - \left\{ a_0/2 + \sum_1^n (a_m \cos mx + b_m \sin mx) \right\} \right|$$

$$\leq \left| \sum_1^n \frac{m}{n} (a_m \cos mx + b_m \sin mx) \right| + \left| \sum_{m=n+1}^\infty \left(1-\frac{1}{n}\right)^m (a_m \cos mx + b_m \sin mx) \right|$$

$$\leq \sqrt{\sum_1^n \frac{m^2}{n^2}(a_m^2+b_m^2)\sum_1^n (\sin^2 mx+\cos^2 mx)}$$

$$+\sqrt{\sum_{n+1}^{\infty}(a_m^2+b_m^2)\sum_{n+1}^{\infty}\left(1-\frac{1}{n}\right)^{2m}(\sin^2 mx+\cos^2 mx)}$$

$$\leq \sqrt{\sum_1^n \frac{m^2}{n}(a_m^2+b_m^2)} + \sqrt{\frac{\left(1-\frac{1}{n}\right)^{2(n+1)}}{1-\left(1-\frac{1}{n}\right)^2}\sum_{n+1}^{\infty}(a_m^2+b_m^2)}$$

$$\leq \sqrt{\sum_1^n \frac{m^2}{n}(a_m^2+b_m^2)} + \sqrt{(n+1)\sum_{n+1}^{\infty}(a_m^2+b_m^2)}.$$

We have already seen that the first radical has the upper bound of its final oscillation no greater than $Q\{f\}$, while we have just shown that the second radical has the upper bound of its final oscillation no greater than $\frac{2}{\pi}Q\{f\}$. Hence the difference between $\sum_1^{\infty}\left(1-\frac{1}{n}\right)^m(a_m \cos mx+b_m \sin mx)$ and the nth partial sum of the Fourier series of $f(x)$ has a final oscillation not exceeding $Q\{f\}(1+2/\pi)$.

From this point we may proceed with an argument precisely parallel to that of §5.

§8. **Quadratic Variation and Trigonometric Interpolation.** Let $f(x)$ be a bounded function of x of period 2π, and let

$$\alpha_{nk}=\frac{2}{2n+1}\sum_1^{2n+1} f\left(\frac{2j\pi}{2n+1}\right)\cos\frac{2jk\pi}{2n+1},$$

$$\beta_{nk}=\frac{2}{2n+1}\sum_1^{2n+1} f\left(\frac{2j\pi}{2n+1}\right)\sin\frac{2jk\pi}{2n+1}.$$

Then

$$f_n(x)=\alpha_{n0}/2+\sum_{k=1}^{n}(\alpha_{nk}\cos kx+\beta_{nk}\sin kx)$$

will be the only trigonometric polynomial of degree n which will

coincide with $f(x)$ for every x of the form $\frac{2k\pi}{2n+1}$. We shall show that if $f(x)$ is of limited quadratic variation, $f_n(x)$ converges everywhere to $(f(x+0)+f(x-0))/2$, and converges uniformly over every interval interior to the one over which $f(x)$ is continuous.

We make use of the auxiliary function

$$\phi_n(x) = a_{n0}/2 + \sum_{k=1}^{n-1} \frac{n-k}{n} (a_{nk} \cos kx + \beta_{nk} \sin kx)$$

$$= \frac{2}{2n+1} \sum_{j=1}^{2n+1} f\left(\frac{2j\pi}{2n+1}\right)$$

$$\times \left[1/2 + \sum_{k=1}^{n-1} \frac{n-k}{n} \left(\cos \frac{2jk\pi}{2n+1} \cos kx - \sin \frac{2jk\pi}{2n+1} \sin kx\right)\right]$$

$$= \frac{2}{n(2n+1)} \sum_{j=1}^{2n+1} f\left(\frac{2j\pi}{2n+1}\right) \left[n/2 + \sum_{k=1}^{n-1} (n-k) \cos k\left(\frac{2j\pi}{2n+1} - x\right)\right]$$

$$= \frac{1}{n(2n+1)} \sum_{j=1}^{2n+1} f\left(\frac{2j\pi}{2n+1}\right) \frac{\sin^2\left[\frac{n}{2}\left(\frac{2j\pi}{2n+1} - x\right)\right]}{\sin^2\left[\frac{1}{2}\left(\frac{2j\pi}{2n+1} - x\right)\right]}. \tag{46}$$

This function, which is a slight modification of one employed by D. Jackson,[6] represents a weighted average of the quantities $f\left(\frac{2j\pi}{2n+1}\right)$, in which those outside any specified neighborhood of x tend to zero weight with increasing n, and in which the weighting of the points to the right of x becomes essentially the same as the weighting of those to the left. Jackson's arguments may be used without essential change, and we see that $\phi_n(x)$ converges to $\frac{1}{2}(f(x+0)+f(x-0))$ wherever this latter quantity exists and converges uniformly to $f(x)$ over any interval interior to one over which $f(x)$ is continuous, provided only that $f(x)$ is bounded, as we have assumed.

[6] A Formula of Trigonometric Interpolation, *Rend. Circ. Math. Palermo* 37 (1914), pp. 371–375.

Now,

$$|\phi_n(x)-f_n(x)| = \left|\frac{k}{n}\sum_1^n(a_{nk}\cos kx+\beta_{nk}\sin kx)\right|$$

$$\leq\sqrt{\frac{1}{n^2}\sum_1^n k^2(a_{nk}^2+\beta_{nk}^2)\sum_1^n(\cos^2 kx+\sin^2 kx)}$$

$$=\sqrt{\frac{1}{n}\sum_1^n k^2(a_{nk}^2+\beta_{nk}^2)}. \tag{47}$$

Now, if u is any number of the form $\frac{2k\pi}{2n+1}$,

$$f\left(u+\frac{2\pi}{2n+1}\right)-f(u)=f_n\left(u+\frac{2\pi}{2n+1}\right)-f_n(u)$$

$$=\sum_1^n\left\{a_{nm}\left[\cos m\left(u+\frac{2\pi}{2n+1}\right)-\cos mu\right]\right.$$

$$\left.+\beta_{nm}\left[\sin m\left(u+\frac{2\pi}{2n+1}\right)-\sin mu\right]\right\}$$

$$=\sum_1^n\left[2\beta_{nm}\cos m\left(u+\frac{\pi}{2n+1}\right)\right.$$

$$\left.-2a_{nm}\sin m\left(u+\frac{\pi}{2n+1}\right)\right]\sin\frac{m\pi}{2n+1}.$$

By formulæ familiar in the theory of trigonometric sums and trigonometric interpolation, we obtain:

$$\sum_1^{2n+1}\left\{f\left(\frac{2(j+1)\pi}{2n+1}\right)-f\left(\frac{2j\pi}{2n+1}\right)\right\}^2$$

$$=\sum_1^n 2(2n+1)(a_{nm}^2+\beta_{nm}^2)\sin^2\frac{m\pi}{2n+1}$$

$$\geq\sum_1^n\frac{8m^2(a_{nm}^2+\beta_{nm}^2)}{2n+1}. \tag{48}$$

If we combine (47) and (48), we see that

$$|\phi_n(x)-f_n(x)|\leq\sqrt{\frac{2n+1}{8n}\sum_1^{2n+1}\left\{f\left(\frac{2(j+1)\pi}{2n+1}\right)-f\left(\frac{2j\pi}{2n+1}\right)\right\}^2}$$

$$\leq\sqrt{\frac{2n+1}{8n}V^{(2)}_{\frac{2\pi}{2n+1}}(f)}. \tag{49}$$

From this point on, the argument exactly parallels that following (38).

§9. **Step-Functions and Quadratic Variation.** Let $f(x)$ be a function of period 2π. Then

$$\left[\frac{1}{b}\int_a^{a+b} f(x)dx-\frac{1}{b}\int_{a+b}^{a+2b} f(x)dx\right]^2=\left[\frac{1}{b}\int_a^{a+b}[f(x)-f(x+b)]dx\right]^2$$
$$\leq \frac{1}{b}\int_a^{a+b}[f(x)-f(x+b)]^2dx$$

Hence

$$\sum_1^n\left\{\frac{n}{2\pi}\int_{u+\frac{2\pi k}{n}}^{u+\frac{2\pi(k+1)}{n}} f(x)dx-\frac{n}{2\pi}\int_{u+\frac{2\pi(k+1)}{n}}^{u+\frac{2\pi(k+2)}{n}} f(x)dx\right\}^2$$
$$\leq\frac{n}{2\pi}\sum_1^n\int_{u+\frac{2\pi k}{n}}^{u+\frac{2\pi(k+1)}{n}}\left[f(x)-f\left(x+\frac{2\pi}{n}\right)\right]^2dx$$
$$=\frac{n}{2\pi}\int_0^{2\pi}\left[f(x)-f\left(x+\frac{2\pi}{n}\right)\right]^2dx.$$

From this we deduce at once that

$$\overline{\lim_{n\to\infty}}\sum_1^n\left\{\frac{n}{2\pi}\int_{u+\frac{2\pi k}{n}}^{u+\frac{2\pi(k+1)}{n}} f(x)dx-\frac{n}{2\pi}\int_{u+\frac{2\pi(k+1)}{n}}^{u+\frac{2\pi(k+2)}{n}} f(x)dx\right\}^2\leq Q(f).$$

Let us put the definition:

$$f_n(x)=\frac{n}{2\pi}\int_{u+\frac{2\pi k}{n}}^{u+\frac{2\pi(k+1)}{n}} f(\xi)d\xi\quad\left[u+\frac{2\pi k}{n}\leq x<u+\frac{2\pi(k+1)}{n}\right].$$

Then

$$\int_0^{2\pi}[f_n(x)-f_{2n}(x)]^2dx$$
$$\leq\frac{\pi}{n}\sum_1^{2n}\left\{\frac{n}{\pi}\int_{u+\frac{\pi k}{n}}^{u+\frac{\pi(k+1)}{n}} f(x)dx-\frac{n}{\pi}\int_{u+\frac{\pi(k+1)}{n}}^{u+\frac{\pi(k+2)}{n}} f(x)dx\right\}^2,$$

since

$$\begin{cases} \left| \dfrac{n}{2\pi} \displaystyle\int_{u+\frac{\pi k}{n}}^{u+\frac{\pi(k+2)}{n}} f(x)dx - \dfrac{n}{\pi} \int_{u+\frac{\pi k}{n}}^{u+\frac{\pi(k+1)}{n}} f(x)dx \right| \\ \leq \left| \dfrac{n}{\pi} \displaystyle\int_{u+\frac{\pi(k+1)}{n}}^{u+\frac{\pi(k+2)}{n}} f(x)dx - \dfrac{n}{\pi} \int_{u+\frac{\pi k}{n}}^{u+\frac{\pi(k+1)}{n}} f(x)dx \right|, \\ \left| \dfrac{n}{2\pi} \displaystyle\int_{u+\frac{\pi k}{n}}^{u+\frac{\pi(k+2)}{n}} f(x)dx - \dfrac{n}{\pi} \int_{u+\frac{\pi k}{n}}^{u+\frac{\pi(k+1)}{n}} f(x)dx \right| \\ \leq \left| \dfrac{n}{\pi} \displaystyle\int_{u+\frac{\pi(k+1)}{n}}^{u+\frac{\pi(k+2)}{n}} f(x)dx - \dfrac{n}{\pi} \int_{u+\frac{\pi k}{n}}^{u+\frac{\pi(k+1)}{n}} f(x)dx \right|. \end{cases}$$

This gives us the result that

$$\overline{\lim_{n\to\infty}} \frac{n}{\pi} \int_0^{2\pi} [f_n(x) - f_{2n}(x)]^2 dx = Q(f).$$

Let $\{w_n(x)\}$ be a set of normal, orthogonal functions over $(0, 2\pi)$, all less than or equal to A in maximum modulus, and such that the partial sum to n terms of the w-series for $f(x)$ is always $f_n(x)$ whenever n is a power of 2. A set of essentially this character has been given by Walsh[7] for the interval $(0, 1)$. Then if $2^m < k < 2^{m+1}$ and $a_j = \int_0^{2\pi} f(\xi) w_j(\xi) d\xi$,

$$\begin{aligned} \left| \sum_{2^m+1}^{k} a_j w_j(x) \right| &\leq A \left| \sum_{2^m+1}^{k} a_j \right| \\ &\leq A \sqrt{(k-2^m) \sum_{2^m+1}^{k} a_j^2} \\ &\leq A \sqrt{2^m \sum_{2^m+1}^{2^{m+1}} a_j^2} \\ &= A \sqrt{2^m \int_0^{2\pi} [f_{2^m}(x) - f_{2^{m+1}}(x)]^2 dx} \\ &\leq A \sqrt{\pi Q(f) + \epsilon} \end{aligned}$$

for any positive ϵ, for a sufficiently large m.

[7] American Journal of Math., V. 45 (1923), pp. 5–24.

Let $f(x)$ be a continuous function. Then over the interval over which it is continuous — possibly only a single point —

$$\max |f(x)-f_n(x)| \to 0.$$

Furthermore, Walsh has shown in the paper already mentioned that the convergence and uniform convergence properties of the w-series of a function over an interval depend only on the properties of the function over any including interval. Hence if $f(x)$ is summable over $(0, 2\pi)$ an argument precisely paralleling that of §5 will show that the w-series of $f(x)$ converges at every point at which $f(x)$ is continuous and uniformly over any interval interior to one over which $f(x)$ is continuous, provided only that $V^{(2)}(f)$ is finite. Moreover, the discussion which Walsh has given of the w-series of functions constant except for a single discontinuity will enable us to show that if $V^{(2)}(f)$ is finite and $f(x)$ is summable, the w-series of f converges at every point which is a binarily rational multiple of π.

NOTE ON OPERATIONAL CALCULUS

By V. Bush

The theory of alternating-current networks in the steady state is in excellent form for engineering use. The theory of transients, on the other hand, is in such involved form as to be unmanageable except in the hands of mathematicians. There is great need for the development of transient analysis into a system of sufficient simplicity to be of use to engineers.

Heaviside made a great stride in this direction in his development of operational methods. They are extremely powerful and simple to use; but they have not come into general use, first because at the time they were developed engineering had not progressed to the point of appreciating or needing them, and second because they are lacking in rigor.

Much work has recently been done along the lines of developing transient analysis; for there is great need at present in practice for more knowledge of transients. This work has largely been away from the operational calculus *per se* rather than toward it. In the attempt to obtain consistency and rigor, mathematicians have largely devoted their efforts to showing the equivalence of the operational method to the integral equation. Such work is valuable and necessary, but it is not directly of service to engineers, for they do not understand the mathematics involved.

The purpose of this note is to attempt to develop a consistent direct operational attack on transients, by utilizing some of Carson's recent results[1] on the equivalent integral equation.

Heaviside's method was briefly as follows: The differential equation between a current and voltage is written in symbolic form

$$i = f(p)e$$

where p is the time differentiator. (We will confine the argument to electrical systems under known applied voltages.) The known voltage e is then supposed to be suddenly applied and afterward constant. That is

$$i = Ef(p)1$$

[1] J. R. Carson, Heaviside Operational Calculus, Bell Technical Journal, November, 1922.

where 1 in Heaviside's notation denotes the discontinuous function which is zero before, and unity after, $t = 0$. This solution is fundamental, for after it is obtained solutions for other applied voltages (*e.g.*, sinusoidal) may be obtained without great labor from Carson's extension formula.[2] The formation of the above equation in a heuristic manner is a simple matter, for all that is necessary is to form the steady state alternating current solution in symbolic form by very familiar methods, and then replace $j\omega$ by p. All that is necessary for a complete solution is then to interpret

$$f(p)1.$$

Heaviside gave one method in his expansion formula by which this can always be thrown into terms of ordinary functions for interpretation. This formula has been since proved by other methods than those used by Heaviside, notably by K. W. Wagner.[3] On systems with distributed constants it is unfortunately extremely cumbersome.

The direct operational method of Heaviside was to transform $F(p)$ algebraically into convenient form, and then to convert term by term, or factor by factor, into terms of ordinary functions by rules that were obtained largely by experience, and by identifying p as the derivative and p^{-1} as the integral of any operand to which they were applied. Heaviside's attack in this manner was full of inconsistencies and contradictions, and this is what drove mathematicians then and since into more rigorous and more complicated methods. For example, on page 40 of E. M. T., Vol. II, he puts

$$\frac{p}{p-a}1=0$$

and obtains a correct result; and on page 294 he puts

$$\frac{p}{p-a}1=\epsilon^{at}$$

and again obtains a correct result. Now by his rules, so far as he states them, either of these formulas may be substantiated.

[2] Carson, Trans. A. I. E. E., 1919, p. 352.
[3] Archiv. f. Elek. IV, 159, 1916.

For we may expand $p/p-a$ in two ways; either

$$-\left(\frac{p}{a}+\frac{p^2}{a^2}+\ \ldots\ \right)1$$

in which case by identifying $p\,1$ with zero we have the first result; or

$$\left(1+\frac{a}{p}+\frac{a^2}{p^2}-\ \ldots\ \right)1$$

in which case by identifying p^{-1} with $\int_0^t$ we obtain the second result. The difficulty in the first case is in identifying an infinite series of discontinuous functions with zero, even when all the discontinuities occur at $t=0$. Thus the direct operational method, while it forms a powerful means of attack, is dangerous in that it leads to contradictions.

There is no need to enlarge upon the benefits of the direct method. Every reader of Heaviside is impressed with the contrast between the simplicity of the operational attack, and the complexity of other methods necessary to obtain the same results.

What we need is a system of definitions and operating rules which will allow of the direct application of the operations involved in the operational solution without introducing inconsistencies. While the substantiating logic may be involved, we must have at hand before operational calculus can meet with general use a body of simple laws which are consistent and sufficient for attack on practical problems.

The most fundamental rule that we must have, if algebraic transformation of operators is to be allowable, is

$$pp^{-1}F(t)=p^{-1}pF(t)=F(t)$$

for any operand $F(t)$ which may occur. Also we must have the other fundamental algebraic rules. It is then easy to show that

$$p^n p^m F(t)=p^m p^n F(t)$$

and that

$$\phi(p)[\psi(p)F(t)]=\psi(p)[\phi(p)F(t)]$$

if ϕ and ψ can be expanded in series.

But if we identify p with d/dt and p^{-1} with $\int_0^t$, then p and p^{-1}

are not in general inverse operators, for example on the continuous operand ϵ^{at}, for

$$p^{-1}[p\epsilon^{at}]=\epsilon^{at}-1.$$

Note for example the circuit containing inductance only, of which the operational solution is

$$i=\frac{1}{L}p^{-1}e.$$

This shows that in this case at least a correct interpretation of p^{-1} is $\int_0^t$ when e is the discontiuuous voltage such that

$$e=0,\ t<0;\ e=\epsilon^{at},\ t>0$$

for then

$$i=\frac{E}{L}\int_0^t \epsilon^{at}dt=\frac{E}{aL}(\epsilon^{at}-1)$$

which is correct. We are thus led to the conclusion that the identification of p with d/dt, and the application of d/dt according to usual rules, under these circumstances is not allowable if algebraic transformation of operators is to be retained. This is a reasonable conclusion, for in the above we do not deal with the continuous function ϵ^{at}; but with a function having this form when $t>0$, and having a discontinuity at $t=0$. We cannot expect to be able arbitrarily to assign definitions to p and p^{-1}, and still have a consistent system. This is the procedure followed by Heaviside, and which led to his difficulties. Fortunately we now have at hand methods by which our rules of operation may be derived rather than assigned. In following more rigorous methods for obtaining our laws of the operational calculus we must take care, however, to retain for the benefit of those who will principally use it the admirable simplicity which is its main asset.

In order to arrive at correct interpretations we fall back upon the equivalent integral equation. Carson has shown (*loc. cit.*) that the operational equation

$$i=f(p)1$$

is equivalent to the integral equation:

$$\frac{f(p)}{p}=\int_{0^-}^{\infty}\epsilon^{-pt}i(t)dt,$$

and he states " The significance of the operational equation and the rules of the Heaviside operational calculus are therefore deducible from the latter equation. The whole problem is thus reduced to the purely mathematical problem of solving the integral equation." The only trouble is that he does not then go ahead and deduce these rules; but prefers to handle the integral equation itself. This is all right for the mathematician but not all right for the engineer. The present note attempts a start on filling in this gap.

In the first place, the fact that the integral equation above is an identity for any value of p immediately justifies the handling of the operator algebraically, splitting into partial fractions, etc., provided interpretation of the several terms obtained is made in accordance with the integral equation. The integral equation thus contains within itself the definitions of any operational forms that may be encountered. Thus interpreted the results are bound to be consistent. Expansion into an infinite series or product may require special discussion.

Let us examine once more the example discussed above. Carson notes,

$$\frac{1}{p-a} = \int_0^\infty \epsilon^{-pt} i(t)dt$$

has as a solution $i = \epsilon^{at}$. This substantiates the operational formula

$$\frac{p}{p-a}1 = \epsilon^{at}. \tag{1}$$

We need not at this point inquire as to the meaning of p in this equation regarded as an operational formula. The formula as a whole is the result desired. The specific meaning of p applied to any operand is necessary only when it stands alone.

Similarly the equation

$$p^{-n-1} = \int_0^\infty \epsilon^{-pt} i(t)dt$$

has as a solution

$$i = \frac{t^n}{n!}$$

which shows that

$$p^{-n}1=\frac{t^n}{n!}. \tag{2}$$

If we assume the validity of operating on both sides of equation (2) by $p-a/p$ we have also immediately

$$\frac{p-a}{p}\epsilon^{at}=1 \tag{3}$$

which gives

$$\left(1-\frac{a}{p}\right)\epsilon^{at}=1$$

or

$$p^{-1}\epsilon^{at}=\frac{1}{a}(\epsilon^{at}-)1$$

thus identifying p^{-1} with $\int_0^t$ in this case at least.

We can obtain (3) in another way. It is the expression for the current flowing in a circuit, the operator for which is $p-a/p$ when a voltage ϵ^{at} is suddenly applied. Now if the voltage $e=0$, $t<0$; $e=1$, $t>0$ is applied to the same circuit the solution will be, using formula (2),

$$i=\frac{p-a}{p}1=\left(1-\frac{a}{p}\right)1=(1-at).$$

In order to get the solution for applied ϵ^{at} we use Carson's extension formula:

$$i=\frac{d}{dt}\int_0^t f(t-y)A(y)dy$$

where $A(t)$ is the solution for unit applied continuous potential, and $f(t)$ is the form of applied potential investigated, in this case ϵ^{at}. This gives

$$\frac{p-a}{p}\epsilon^{at}=\frac{d}{dt}\int_0^t \epsilon^{at}\epsilon^{-ay}(1-ay)dy$$

$$=\frac{1}{a}\frac{d}{dt}\epsilon^{at}[1-\epsilon^{-at}+\epsilon^{-at}(1+at)-1]=1.$$

So we may take as one of our rules

$$\frac{p-a}{p}\epsilon^{at}=1. \tag{3}$$

Operate with p on this equation and we have another important formula

$$p\epsilon^{at}=a\epsilon^{at}+p1. \tag{4}$$

Now the $p1$ is an essential part of this formula. It cannot be put equal to zero without leading to contradictions. On the other hand its reiteration avoids just such difficulties. It may be legitimate in a final expression to set $p1=0$ when $t>0$, but it is illegitimate to drop it out before the final interpretation, or even then if it occurs in an infinite series of such terms. It is the retention of this function $p1$ that is the principal point in the present note. It is useless to inquire as to the physical meaning of $p1$ as it occurs during an analysis, just as useless as it would be to attempt to physically interpret $\sqrt{-1}$ when it occurs under similar conditions. The expression $p1$ is still an operational expression, subject to operational rules and transformations, but surely not to be discarded in the midst of an argument.

Note that we now have

$$p^{-1}[p\epsilon^{at}]=p^{-1}[a\epsilon^{at}+p1]=\epsilon^{at}-1+1=\epsilon^{at}$$

so that p and p^{-1} are inverse operators when applied to ϵ^{at}.

We may obtain

$$p^{-n}\epsilon^{at}=\frac{1}{a^n}\left[\epsilon^{at}-1-at-\frac{(at)^2}{2}-\quad\cdots\quad\frac{(at)^{n-1}}{(n-1)\colon}\right] \tag{5}$$

from the extension theorem, using (2), so that

$$p^{-n}\epsilon^{at}=\frac{1}{n\colon}\frac{d}{dt}\int_0^t\epsilon^{at}\epsilon^{-ay}y^n dy.$$

Successive integrations by parts on this expression give the result (5). Note that this again corresponds to $p^{-1}=\int_0^t$.

Now from (4)

$$p1=(p-a)\epsilon^{at}. \tag{6}$$

Operating on this with p we have

$$\begin{aligned} p^2 1&=p(p-a)\epsilon^{at}\\ &\cdots\cdots\cdots\\ p^n 1&=p^{n-1}(p-a)\epsilon^{at}. \end{aligned} \tag{7}$$

A similar expression

$$p^{-n}1 = p^{-n-1}(p-a)\epsilon^{at} \tag{8}$$

readily follows from (2). From (4)

$$p^2\epsilon^{at} = a^2\epsilon^{at} + ap1 + p^21$$

and in general

$$p^n\epsilon^{at} = a^n\epsilon^{at} + p^n1 + ap^{n-1}1 + a^2p^{n-2}1 + \ . \ . \ . \ a^{n-1}p1. \tag{9}$$

Using (7) we may put this in the form

$$p^n\epsilon^{at} = a^n\epsilon^{at} + (p-a)(a^{n-1} + a^{n-2}p + \ . \ . \ . \ p^{n-1})\epsilon^{at}.$$

The operational formula

$$\frac{p}{\sqrt{p^2+a^2}}1 = J_0(jat) \tag{10}$$

may be established by expanding in series and using only $p^{-n}1 = t^n/n!$ [4]

Similarly we can get [5]

$$\left(\frac{p}{p+2a}\right)^{\frac{1}{2}}1 = \epsilon^{-at}J_0(jat). \tag{11}$$

Note that

$$\left(\frac{p}{p+2a}\right)^{\frac{1}{2}}1 = \epsilon^{-at}\frac{p}{(p^2-a^2)^{\frac{1}{2}}}1$$

suggests the general transformation formula

$$f(p)1 = \epsilon^{-at}f(p-a)\,\frac{p}{p-a}1 \tag{12}$$

although it does not prove it in general. This transformation is very useful in removing the attenuation factor from a wave solution before examining it operationally. The part remaining after splitting off the factor ϵ^{-at} in this manner is the unattenuated wave, and is ordinarily simplified by the transformation. This formula is usually equivalent in results to one used by Heaviside for change of operand; but his breaks down under the present interpretation, and the corrected form of his formula is cumbersome. Note also that we do not have in general $\phi(p)\epsilon^{at}$ equal to $\phi(a)\epsilon^{at}$, as he often assumed.

[4] See notes by M. S. Vallarta, M. I. T., Course 6.51, 1923, p. 17, pt. I.
[5] Same notes, pt. II, p. 12.

In each of these cases terms of the form $p^n 1$ enter. When they appear in a solution they can be avoided by changing operand. Thus for example we have

$$\frac{p}{p-a}1 = -\left(\frac{p}{a}+\frac{p^2}{a^2}+ \ . \ . \ .\right)1.$$

Put for 1 its equivalent $\frac{p-a}{p}\epsilon^{at}$ and this becomes

$$-\left(\frac{1}{a}+\frac{p}{a^2}+ \ . \ . \ .\right)(p-a)\epsilon^{at}$$

$$= \left[\left(1+\frac{p}{a}+\frac{p^2}{a^2}+ \ . \ . \ .\right)-\left(\frac{p}{a}+\frac{p^2}{a^2}+ \ . \ . \ .\right)\right]\epsilon^{at} = \epsilon^{at}.$$

The introduction of fractional powers of the operator p introduces no new difficulties. I have shown from Cauchy's integral formula that

$$p^{1/2}1 = (\pi t)^{-1/2} \tag{13}$$

when 1 has its discontinuous meaning.[6]

This also follows from Carson's integral equation,[7] using the identity

$$p^{-1/2} = \int_0^\infty \epsilon^{-pt}(\pi t)^{-1/2}dt.$$

We cannot, however, obtain $p^{3/2}1$ from (13) directly.

Let us next examine

$$\frac{a}{p-a}t^{-1/2}.$$

If this is interpreted in the usual manner, using $p^{-1} = \int_0^t$, we have

$$\frac{a}{p-a}t^{-1/2} = a\epsilon^{at}\int_0^t \epsilon^{-at}t^{-1/2}dt \tag{14}$$

and this may be expanded by integrations by parts into

$$\frac{a}{p-a}t^{-1/2} = \sqrt{a\pi}\epsilon^{at} - t^{-1/2}\left[1-\frac{1}{2}\left(\frac{1}{at}\right)+\frac{1}{2}\frac{3}{2}\left(\frac{1}{at}\right)^2 \ . \ . \ .\right]. \tag{15}$$

In fact we may obtain, when $n > -1$

$$\frac{a}{p-a}t^n = a\epsilon^{at}\int_0^t \epsilon^{-at}t^n dt. \tag{16}$$

[6] Discussion, Trans, A. I. E. E., 1919, p. 469.
[7] Carson, Bell Jour., November, 1922, p. 51.

We may obtain these results directly. Thus to establish (14) we have the extension formula previously used

$$\frac{a}{p-a}t^{-1/2}=\frac{d}{dt}\int_0^t(t-y)^{-1/2}(\epsilon^{ay}-1)dy.$$

Splitting into two terms and integrating by parts gives (14) directly.

It is interesting to examine one of Heaviside's solutions in view of the new operational rules developed above. In E. M. T., Vol. II, p. 40, he treats the case of an infinite cable with a terminal condenser. His operational solution comes into the form:

$$V_0=(1-qa+q^2a^2-q^3a^3+\ \ .\ \ .\ \ .)qal$$

where q is $\sqrt{RCp}$. He then throws out even powers of q, or positive integral powers of p, operating on 1. He says this is not as simple a matter as it looks, and it surely is not. This is equivalent to using $(p/p-a)1$ equal to zero. He next uses

$$p^{1/2}1=(\pi t)^{-1/2}$$

which is legitimate, but then he uses

$$pt^{-1/2}=-\tfrac{1}{2}t^{-3/2}\text{ etc.}$$

which is not. The striking thing is that he comes out with the correct result, although, as Carson notes, exactly the same procedure in other cases, such as a terminal inductance, leads to absurdities. Now if we use our formula (1)

$$\frac{p}{p-a}1=\epsilon^{at}$$

and if we use (15) we introduce two new terms in ϵ^{at}. Upon following through the algebra again we find that these terms will cancel and give us Heaviside's correct result. More important we now have a correct guide for use in other cases.

We have as yet no formula for $pt^{-1/2}$ etc. If we write (15) with $a=1$ in the form

$$(1+p+p^2+\ \ .\ \ .\ \ .\ \ )t^{-1/2}$$

$$=\left(t^{-1/2}-\frac{1}{2}t^{-3/2}+\frac{1}{2}\frac{3}{2}t^{-5/2}.\ \ .\ \ .\right)-\sqrt{\pi}\epsilon^{at},$$

and if we insert for ϵ^t its value from (1)

$$-(p+p^2+\ \ldots\ .)1,$$

we can write

$$(p+p^2+\ \ldots\ .)t^{-1/2}$$

$$=\sqrt{\pi}(p+p^2+\ \ldots\ .)1+\left(-\frac{1}{2}t^{-3/2}+\frac{1}{2}\,\frac{3}{2}t^{-5/2}\ \ldots\ .\right)$$

This is satisfied by

$$\left.\begin{aligned} pt^{-1/2}&=-\tfrac{1}{2}t^{-3/2}+\sqrt{\pi}p1\\ p^2t^{-1/2}&=\frac{1}{2}\,\frac{3}{2}t^{-5/2}+\sqrt{\pi}p^21\\ &\ldots\ldots\ldots\ldots\\ p^nt^{-1/2}&=\frac{1.3.5\ \ldots\ (2n-1)}{2^n}t^{-n-1/2}+\sqrt{\pi}p^n1 \end{aligned}\right\}\quad(17)$$

and this set of values is consistent with

$$p^n(p^mt^{-1/2})=p^{m+n}t^{-1/2}$$

The rules so far developed will be collected together for convenient reference.

$$\frac{p}{p-a}\,1=\epsilon^{at} \qquad (1)$$

$$\frac{p-a}{p}\,\epsilon^{at}=1 \qquad (2)$$

$$p^{-n}1=t^n/n!$$

$$p\epsilon^{at}=a\epsilon^{at}+p1 \qquad (3)$$

$$p^{-n}\epsilon^{at}=\frac{1}{a^n}\left(\epsilon^{at}-1-at-\frac{(at)^2}{2!}-\ \ldots\ .\frac{(at)^{n-1}}{(n-1)!}\right) \qquad (4)$$

$$p1=(p-a)\epsilon^{at} \qquad (5)$$

$$p^n1=p^{n-1}(p-a)\epsilon^{at} \qquad (6)$$

$$p^{-n}1=p^{-n-1}(p-a)\epsilon^{at} \qquad (7)$$

$$p^n\epsilon^{at}=a^n\epsilon^{at}+(p-a)(a^{n-1}+a^{n-2}p+\ \ldots\ .\ p^{n-1})\epsilon^{at}$$

$$p^n\epsilon^{at}=a^n\epsilon^{at}+p^n1+ap^{n-1}1+\ \ldots\ .\ a^{n-1}p1 \qquad (8)$$

$$\frac{p}{\sqrt{p^2-a^2}}1 = J_0(jat) \tag{9}$$

$$\left(\frac{p}{p+2a}\right)^{1/2}1 = \epsilon^{-at}J_0(jat) \tag{10}$$

$$f(p)1 = \epsilon^{-at}f(p-a)\frac{p}{p-a}1 \tag{12}$$

$$p^{1/2}1 = (\pi t)^{-1/2} \tag{13}$$

$$\frac{a}{p-a}t_n = a\epsilon^{at}\int_0^t \epsilon^{-at}t^n dt \tag{16}$$

$$n > -1$$

$$\frac{a}{p-a}t^{-1/2} = \sqrt{a\pi}\epsilon^{at} - t^{-1/2}\left(1 - \frac{1}{2}\left(\frac{1}{at}\right) + \frac{1}{2}\frac{3}{2}\left(\frac{1}{at}\right)^2 - \ldots\right) \tag{15}$$

$$p^n t^{-1/2} = \frac{1.3.5\ \ldots\ (2n-1)}{2}t^{-n-1/2} + \sqrt{\pi}p^n 1 \tag{17}$$

Not all of these have been completely proved. Other rules may be developed as needed. A collection of such operational rules bears the same relation to operational calculus that a table of integrals does to integral calculus. The above table may readily be extended by further transformations, falling back upon the integral equation where necessary, in the same manner that we occasionally have recourse to fundamental definitions in building up a table of integrals.

The method of solving a practical transient problem by the use of the operational calculus is as follows: We first set up the operational solution, usually by analogy with the steady state periodic solution which is well known. This operational solution is then transformed by any algebraic processes, and by the use of operational transformation formulas from the tables, until it is in such a form as to be recognized and interpreted. This final form may be an expression of the current as a time function, or it may be still an operational expression, the nature of which is known from the tables. The utility of this latter procedure should be emphasized. We deal in transient study with discontinuous functions. The expressions for these in ordinary terms, even as Fourier series, are often exceedingly cumbersome. The natural

manner of expressing a discontinuous function is by means of an operational expression, and great simplicity of form may be attained by so doing. Such forms when properly substantiated and shown to be unique, will be just as precise and complete statements of the functions they represent as any series expansion into which they may be converted. The proper procedure is to retain the operational forms wherever possible, and interpret them as they stand.

Of course the method applies equally well to mechanical and electrical problems, and to know applied currents or velocities as well as to known applied potentials or forces.

NOTE ON THE QUANTIZATION OF NON-CONDITIONED-PERIODIC SYSTEMS

By M. S. VALLARTA

1. Among the most interesting problems which have presented themselves in connection with the development of Bohr's theory of the atom, that of the quantization of unconditioned-periodic dynamical systems is of paramount importance, standing as it does in the way of further analytical extensions of the quantum theory of stationary paths. The following remarks, some of which have already found expression elsewhere, while failing to supply a final answer to this question, seek to contribute a few guiding ideas to the elucidation of the problem. We shall first examine the problem from the standpoint of ordinary dynamics.

Consider a holonomic conservative reversible dynamical system of n degrees of freedom specified by the system of canonical equations:

$$\dot{p}_k = -\frac{\partial H}{\partial q_k}, \qquad \dot{q} = \frac{\partial H}{\partial p_k} \qquad k = 1, 2 \ldots n$$

H, the Hamiltonian function, is a function of p and q only. Assume a transformation $p, q \rightarrow P, Q$ specified by

$$p_k = p_k(P_1 \ldots P_n, Q_1 \ldots Q_n), \qquad q_k = q_k(P_1 \ldots P_n, Q_1 \ldots Q_n)$$
$$k = 1, 2 \ldots n$$

where P, Q satisfy the relations

$$\begin{aligned} &[P_i, P_k] = 0 \qquad [Q_i, Q_k] = 0 \qquad\qquad i, k = 1, 2 \ldots n \\ &[Q_i, P_k] = 0 \quad , \quad i, k = 1, 2 \ldots n \qquad i \neq k \\ &[Q_i, P_i] = 1 \qquad\quad i = 1, 2 \ldots n \end{aligned}$$

in which $[u, v] = \sum_{r=1}^{n} \left(\frac{\partial q_r}{\partial u} \frac{\partial p_r}{\partial v} - \frac{\partial p_r}{\partial u} \frac{\partial q_r}{\partial v} \right)$ are the Lagrangian bracket expressions. The transformation $p, q \rightarrow P, Q$ is then a contact-transformation,[1] and it is shown that both the Hamilton function and the canonic equations are invariant for such a transformation.

[1] Whittaker, Analytical Dynamics, p. 298, Cambridge, 1917.

Let W be a function of the variables q, Q and perform the contact-transformation

$$P_k = -\frac{\partial W}{\partial Q_k} \qquad p_k = \frac{\partial W}{\partial q_k} \qquad k = 1, 2, \ldots n$$

and let W be an integral of the Hamilton-Jacobi partial differential equation,

$$-E + H\left(q, \frac{\partial W}{\partial q}\right) = 0$$

E being the energy-constant. Jacobi's well-known theorem is then to the effect that, if a complete solution of Hamilton-Jacobi's equation containing n arbitrary constants $\alpha_1 \ldots \alpha_n$ can be found, then

$$\beta_k = -\frac{\partial W}{\partial \alpha_k} \qquad \rho_k = \frac{\partial W}{\partial q_k} \qquad k = 1, 2, \ldots n$$

give the solution of the dynamical problem.

2. The integration of the Hamilton-Jacobi equation by separation of variables, *i.e.*, by expressing W as a sum of functions $\sum_{i=1}^{n} W_i(q_i)$, each of which depends on but one variable only, plays, as is well-known, a fundamental part in the analytical investigation of stationary orbits, as determined by the Sommerfeld-Wilson conditions. Under the assumption that the Hamilton-Jacobi equation is integrable in only one way by separation of variables, the Hamilton function depends on the I's only, where I_k is the action variable canonically conjugate to p_k, q_k, and the system is uniquely quantized both as regards paths and energy.

As far as the writer is aware, a complete investigation of the conditions under which Hamilton-Jacobi's equation is integrable by separation of variables has not yet been undertaken, and great are the difficulties to be overcome if a complete discussion of the general dynamic system of n degrees of freedom is aimed at. The problem may be concisely stated as follows: Given a dynamical system, can we establish whether it leads to a Hamilton-Jacobi equation which can be integrated by separation of variables?

A fundamental step towards the solution of this question was

made by T. Levi-Civita.[2] We notice in the first place that the hypothesis that W be of the form

$$W=\sum_{i=1}^{n} W_i(q_i)$$

is equivalent to $\frac{dp_i}{dq_j}=0$ for $i\neq j$. If we differentiate our Hamilton-Jacobi equation with respect to q_i we have

$$\frac{\partial H}{\partial q_i}+\frac{\partial H}{\partial p_i}\frac{\partial p_i}{\partial q_i}=0$$

but since none of the derivatives $\frac{\partial H}{\partial p_i}$ can vanish identically we may write out the above as:

$$\frac{\partial p_i}{\partial q_i}=-\frac{\frac{\partial H}{\partial q_i}}{\frac{\partial H}{\partial p_i}}$$

hence all the derivatives of the p's with respect to the q's are now defined. In order that these and W may exist, the condition of integrability

$$\frac{d}{dq_j}\left(\frac{\frac{\partial H}{\partial q_i}}{\frac{\partial H}{\partial p_i}}\right)=0$$

must be satisfied.

Since all the derivatives of the p's with respect to the q's are defined, at most their initial values remain arbitrary, which besides must be so if W is a complete integral of Hamilton-Jacobi's equation. It follows that the necessary and sufficient condition that the variables be separable is that the integrability condition be identically satisfied with respect to the $2n$ variables p_k and q_k. We therefore conclude with Levi-Civita:

The Hamilton-Jacobi equation is integrable by separation of variables when, and only when, the Hamiltonian function H

[2] Mathematische Annalen. Vol. 59, p. 383, 1904.

satisfies all $n(n-1)/2$ equations of the second order (obtained by expansion of the last equation):

$$\frac{\partial H}{\partial p_i}\frac{\partial H}{\partial p_j}\frac{\partial^2 H}{\partial p_j \partial q_j} - \frac{\partial H}{\partial p_i}\frac{\partial H}{\partial q_j}\frac{\partial^2 H}{\partial q_i \partial p_j} - \frac{\partial H}{\partial q_i}\frac{\partial H}{\partial p_j}\frac{\partial^2 H}{\partial p_i \partial q_j} + \frac{\partial H}{\partial q_i}\frac{\partial H}{\partial q_j}\frac{\partial^2 H}{\partial p_i \partial p_j} = 0 \qquad (1)$$

$i \neq j$

Levi-Civita has furthermore shown what conditions are to be satisfied by the kinetic and the potential energy of a system in order that the Hamilton-Jacobi equation be separable. His treatment is based upon the well-known theorem[3] that the determination of the motion of a particle moving in a conservative force-field is reducible to the problem of finding the geodesics in an n-dimensional space with given line element. The problem is thus reduced to a purely geometric investigation and it is shown that if a dynamical system having a Hamiltonian function $H = E$ is integrable by separation of variables, the same is true of the equation which defines the geodesics. We thus have that in order to classify the dynamical systems which are soluble by separation of variables we must

(1) Characterize the kinetic energy, or what amounts to the same, the linear element $ds^2 = \sum_{i,k=1}^{n} a_{ik} dx_i dx_k$ which corresponds to it.

(2) For each of the cases determined under (1), investigate which force-fields can be applied.

A discussion covering (1) has been given by Stäckel[4] and Levi-Civita for the case $n = 2$ and by dall'Acqua[5] for the case $n = 3$. In a special case, the system of n degrees of freedom has also been discussed by Levi-Civita (l.c.). The complete investigation as regards (1) and (2) has, to the writer's knowledge, never been undertaken.

The range of physical questions to which the quantum theory is applicable is such that no investigation can be considered

[3] For ex. Whittaker, l c., p. 254, also Painlevé.
[4] Mathematische Annalen. Vol. 42, p. 537, 1893.
[5] Mathematische Annalen. Vol. 66, p. 398, 1909.

too general for the purpose. The quantum theory and the separation of variables are so intimately connected that perhaps it is safe to say that the treatment of the former depends upon the solution of the latter.

3. We shall now define a conditioned periodic system as one whose Hamiltonian function satisfies the Levi-Civita conditions (1) and a non-conditioned-periodic system as one for which these conditions are not satisfied. All attempts to quantize the latter are based on a reduction to a conditioned periodic motion, because it is only for such motions that the quantum conditions acquire a meaning. In recent papers P. S. Epstein[6] has utilized Delaunay's[7] perturbation theory to devise a method of approximating to the *general* solution of the canonic equations governing a non-conditioned-periodic system of n degrees of freedom by means of a conditioned periodic motion expressed as an n-fold Fourier series. His method is essentially as follows: Suppose in general that the Hamilton-Jacobi equation of the system is non-integrable by separation of variables. We determine a conditioned periodic motion of Hamiltonian function H_{p1} (p, q) which comes as near as possible to the actual H and split the latter into two parts

$$H = H_{p1} + C_1$$

where C is a function of momenta and cordinate expressing the "perturbation." A contact-transformation from variables p, q to variables α, w defined by

$$W(q, \alpha) \quad p_k = \frac{\partial W}{\partial q_k} \quad w_k = \frac{\partial W}{\partial \alpha_k} \quad k = 1, 2 \ . \ . \ . \ n$$

may now be made. Now, *if* — and this assumption must be clearly emphasized in view of later developments — the perturbation function C is everywhere regular within the region considered, C may be expanded in an n-fold Fourier series and since H_{p1} is now a function of α's only, we have

$$H = H_{p1}(\alpha) + \sum_{m_1 \ . \ . \ . \ m_n = 1}^{n} a_{m_1} a_{m_2} \ . \ . \ . \ a_{m_n} \begin{matrix}\cos \\ \sin\end{matrix} (m_1 w_1 + m_2 w_2 + \ . \ . \ . \ m_n w_n)$$

[6] Zeitschrift für Physik. Vol. 8, pp. 211, 305, 1922; Vol. 9, p. 92, 1922.
[7] Théorie du movement de la lune. Paris, 1860.

where the a's are functions of the momenta α. It is seen, therefore, that the " perturbed " motion itself may be approximated to as closely as desired by conditioned periodic motions. Hence the Sommerfeld-Wilson conditions still are able to yield, to any desired degree of approximation on the basis of Epstein's method, the stationary orbits. Thus " sharp " quantization — in a physical sense — of non-conditioned-periodic systems is always possible.

Bohr[8] and Kramers[9] have also given methods of quantization which depend on the investigation of the perturbation. The essential feature is the introduction of angle variables to describe the perturbed system. Their method, which may be looked upon as a first approximation, was extended by Born and Pauli[10] who devised a *formal* procedure by means of which the approximation to the actual motion by means of the conditioned periodic motion may be formally pushed as far as desired. Start from a conditioned periodic system — therefore rigorously integrable by separation of variables — and let it become non-conditioned-periodic by the influence of a perturbing agency. Let W_k^0 be the angle variables of the initial motion, I_k^0 the canonically conjugate action variables and $H_0\,(I_k^0)$ the Hamiltonian function. Now expand the Hamiltonian function H of the perturbed motion in a power series in the perturbing (small) parameter λ:

$$H = H_0 + \lambda H_1 + \lambda^2 H_2 + \ . \ . \ .^{11}$$

The problem is then to determine a function $F(I_k, W_k^0)$ having the property

$$F = I_k W_k^0 + \text{periodic function of } W_k^0 \text{ of period 1}$$

which makes

$$H = H_0\left(\frac{\partial F}{\partial W_k^0}\right) + \lambda H_1\left(\frac{\partial F}{\partial W_k^0}, W_k^0\right) + \lambda^2 H_2\left(\frac{\partial F}{\partial W_k^0}, W_k^0\right) + \ . \ . \ .$$

[8] Mém. de l'Acad. Roy. des Sci. et Lettres du Danemark. Vol. 4, No. 1, fasc. 1, 2,; 1918.
[9] Zeitschrift für Physik. Vol. 3, p. 201, 1920.
[10] Zeitschrift für Physik. Vol. 10, p. 137, 1922.
[11] For the degenerate case see besides l.c. Born and Heisenberg Zeitschrift für Physik, Vol. 14, p. 48, 1923.

independent of the angle variables W_k^0. We shall not go into this any further, but refer to the original memoir for details. It is seen that again the essential step is the approximation to the actual motion by a conditioned periodic motion. Just as in the case of Epstein's method it would appear that sharp quantization is possible.

4. A closer examination of the matter, however, shows that doubts as to the correctness of the last conclusion may be rightly entertained. Born and Pauli have shown on the basis of known theorems of Bruns[12] and Poincaré that, however small λ may be, the process outlined above is *not* in general convergent. It follows that, in those cases where the *general* solution does converge, it cannot *in general* approach either Born-Pauli's approximate solution or Epstein's, continuously. In particular, it is not in general true that a Delaunay transformation can actually result in decreasing the perturbation function. Still another argument may be brought up: In a recent memoir, Smekal[13] has pointed out that, according to a theorem of Herglotz,[14] if a continuous $2n$-dimensional region of phase-space exists where an n-fold Fourier Series converges everywhere uniformly, then a contact-transformation can be always found which permits obtaining the general solutions directly; since then all the coördinates of the system are cyclical, it is separable and conditioned periodic, which contradicts the initial assumption. The theorem may therefore be stated: No general solution of a non-conditioned-periodic system can be expanded in a uniformly convergent Fourier series.

We conclude, with Smekal and others, that should n-fold periodic solutions exist at all for non-conditioned-periodic systems, they must be *particular* solutions obtained by ascribing values to some of the arbitrary constants in the general integrals, which in turn depend on certain parameters. From the point of view of such fundamental concepts in the development of the quantum theory as the adiabatic invariance of the quantized dynamical

[12] See for instance Charlier, Mechanik des Himmels. Vol. 2, p. 307.
[13] A. Smekal, Zeitschrift für Physik. Vol. 11, p. 294, 1922.
[14] Quoted by Smekal, l.c., p. 295. It has not been possible to consult the original.

solutions and the Principle of Correspondence, which Bohr[15] and Kramers[16] rightly place at the basis of the theory of stationary orbits, such particular solutions lead to renewed difficulties.

5. Assume that the parameters on which the arbitrary constants depend undergo an adiabatic transformation. Then the result is not only that the new initial conditions do not correspond to any of the particular n-fold periodic solutions (assuming that the initial conditions vary continuously) but also that in general the new solution is not at all useful for the purposes of the quantum theory of stationary states, because it is an asymptotic or "unstable" solution, that is, one in which several or all the variables involved increase with time without limit. Nor can such solutions be brought into agreement with the fundamental conceptions of the Principle of Correspondence. For in order to carry out the passage to the limit by means of Bohr's frequency relation, the Hamiltonian function must be a continuous function of the n action variables. The conclusion is therefore forced on us that it cannot be expected that the methods worked out by Epstein or by Born-Pauli will *in general* enable us to calculate, even approximately, the quantized states of non-conditioned-periodic systems. If, on the other hand, we were to look for the quantized states among the general, unknown, solutions, we have neither a quantization rule nor any proof of the existence of adiabatic invariants. Smekal has, however, pointed out an exception to this last statement for microcanonic (quasi-ergodic[17]) statistical systems in which a single adiabatic invariant (the volume bounded by the energy-surface) is known to exist. Since normal mechanical systems are quasi-ergodic,[18] the argument applies equally well to all such systems, which must, therefore, be susceptible of sharp quantization.

[15] N. Bohr, Zeitschrift für Physik. Vol. 13, p. 117, 1923.
[16] H. A. Kramers, Zeitschrift für Physik. Vol. 13, p. 312, 1923.
[17] A quasi-ergodic system is one whose path in phase-space lies wholly within two energy-surfaces which differ by an infinitesimal energy increase. This is the same as the microcanonical ensemble in Gibbs' sense. The impossibility of the existence of a "strictly" microcanonical (ergodic) system, *i.e.*, one in which the path in phase-space lies wholly upon a given energy-surface, was established by Rosenthal and Plancherel, cf. Rosenthal, Annalen der Physik, Vol. 42, p. 796, 1913, Plancherel, ibid., p. 1061.
[18] E. Fermi, Nuovo Cimento, Vol. 69, p. 267, 1923.

6. Suppose, on the other hand, that quantized solutions are to be looked for among *other* particular solutions. Then, as was first pointed out by Smekal, we have again the possibility of selecting solutions of conditioned periodic character which converge everywhere uniformly within a continuous multi-dimensional region and are obtained by ascribing definite functional relations among the different integrals.[19] It is clear that such solutions have less than n independent periods and hence correspond to an s-space of lower dimensionality than n. The treatment of non-conditioned-periodic systems would then be similar to that of ordinary degenerate problems. Simple-periodic particular solutions were studied by Poincaré.[20] Multiple-periodic solutions may be derived from these in certain special cases by the method of characteristic exponents.[21]

It may be carefully emphasized that such s-fold periodic solutions offer no difficulty from the point of view of the adiabatic invariance of quantized orbits, provided the adiabatic transformation is such as to leave the functional relations invariant. The system as a whole behaves as a conditioned periodic system of s degrees of freedom ($s < n$). Likewise, there is no difficulty in the application of the Principle of Correspondence. "Sharp" quantization is thus always possible and *all* atomic dynamical systems are fundamentally integrable by separation of variables. With regard to parametric changes which do not leave our functional relations invariant, it may be assumed that such parameters do not vary in nature. Provisionally, the postulate may be looked upon in a light similar to that of the existence of non-radiating orbits. Likewise, it should be noticed that our point of view, that the system behaves as an s-fold periodic one, does not at all imply that the number of quantum numbers which characterize a given quantized path is less than the number of degrees of freedom.[22] It only implies that the number (s) of quantum numbers to which correspond integral values 0, 1, 2,

[19] T. Levi-Civita. C. R. 131; 170, 236; 1900.
[20] Poincaré. Acta Mathematica 13, 248, 1890, see also Charlier, Mechanik des Himmels, Vol. 2, p. 187.
[21] See for ex. Whittaker, l.c., p. 400.
[22] Born and Heisenberg, l.c. in our Note (11); p. 52, footnote 2 also, Smekal, Zeitschrift für Physik. Vol. 15, p. 60, 1923, footnote 1.

. . . is *less* than the number of degrees of freedom (n). The other ($n-s$) quantum numbers must therefore be held at fixed constant values. Hence the system as a whole is characterized by n quantum numbers, of which s are given values 0, 1, 2, 3, . . . and $n-s$ stay at any fixed integral values. Spectroscopically, this would mean that the degree of multiplicity of a term is less than n.

In support of this view, it may be mentioned that all the spectroscopic experimental evidence available at present shows that the number of " optical " degrees of freedom is much less than the number of dynamical degrees of freedom calculated from the number of particles of which the atomic system is made up (*cf.* Smekal, l.c.). Also, as far as the writer is aware, all experimental evidence points towards the existence of " sharp " spectral lines only.

It may be remarked that we do not take at all into consideration here the difficulties which have led some writers to conclude the " failure of classical mechanics " and the existence of " unmechanical " orbits in atomic systems (*cf.* for example l.c. **15, 16**). Our views, however, do not necessarily come in conflict with such conclusions, because many of the theories expressed above can be transferred *mutatis mutandi* to " non-classic " mechanical orbits, provided only that these satisfy differential equations of the second order.

Research Laboratory of Theoretical Physics
January, 1924

A CONTRIBUTION TO THE GENERALIZATION OF A DETERMINANTAL THEOREM OF FROBENIUS

By LEPINE HALL RICE

1. In 1894, Frobenius[1] gave a theorem purely on determinants, which Muir[2] in 1922 generalized. The theorem and the generalization were stated only for ordinary or 2-way determinants. Lecat,[3] in 1923, made the extension to any number of dimensions, but this was not his only advance, for he announced that the relations obtained by Muir held good for a distribution of elements in the matrix of the given determinant not restricted to the forms contemplated by Muir, whether in the 2-way or in the p-way[4] case. In this and in other important respects Lecat's paper is of immediate concern to one who uses only 2-way determinants.[5]

Each of Muir's relations involves a sum of determinants whose matrices are formed by altering certain elements in the matrix of a given determinant Δ. The relations first given are produced by changing elements of Δ into zeros. Muir then shows how arbitrary quantities can be made to take the place of some of these zeros in one type of matrix, and Lecat formulates this modification for p dimensions. It is one object of the present paper to give rules for the introduction of arbitrary quantities in the case of the most general distribution of elements, both in the way contemplated by Muir and Lecat and in a second way. It is a further object to take a point of view of the relations thus

[1] Ueber die Elementartheiler der Determinanten. Sitzungsb . . . Akad. d. Wiss, (Berlin), 1894, pp. 31–44, Sec. 4.

[2] Note on a theorem of Frobenius connected with invariant-factors. Proc. R. Soc. Edinburgh, XLII (1921–1922), pp. 342–347.

[3] Généralisations et modifications d'un théorème de Frobenius. Ann. de la Soc. scientifique de Bruxelles, XLII, prem. pt., pp. 322–329.

[4] We prefer to reserve the letter n, as in previous papers, for the *order* of the determinant or matrix, thus securing uniformity when we have occasion to pass from two dimensions to more, or vice versa. Besides, the n dimensions of geometry seem much more closely related to the order than to the number of "dimensions" of a general determinant.

[5] For the benefit of such a one, we state some matters, in the following pages, in more detail and with more definitions than a person familiar with the literature of determinants of many dimensions requires.

far announced from which the proof of them appears a very simple and almost obvious matter. The methods of proof that will be followed have been effectively used many times by writers in the past. They are elemental, not involved, not dependent on the presence of such regular formations in matrices as are required, for example, in order that Laplace's or Cayley's expansion may be applied. These expansions look like powerful instruments of research and often they are; but on the other hand there are times when we rest content with a theorem to which they lead us instead of searching for another and simpler mode of proof of the same theorem in the hope that some condition in it may thus be shown to be not inherent but imported with the first proof.

2. The theorem of Frobenius concerns a general 2-way determinant Δ of order n, in whose matrix — let us say in the upper left-hand or "leading" corner — a rectangular array B of r rows and at least r columns is marked off. All possible products $D_kD'_k$ are now formed, where D_k is a minor of order k lying in B, and D'_k is its cofactor in Δ, k taking all values from 1 to r. Also a determinant Δ_C is formed from Δ by changing into zeros all the elements in B, leaving intact the remaining elements of Δ, whose locus we may denote by C. Then the theorem is:

$$\Delta - \Sigma D_1D'_1 + \Sigma D_2D'_2 - \ . \ . \ . + (-1)^r \Sigma D_rD'_r = \Delta_C \qquad (1)$$

The proof given by Frobenius was by differentiating the left member with respect to any element of B and showing that the derivative vanished, whence the left member must be independent of the elements of B and thus must reduce to Δ_C.

In this proof the derivative involves sums of products of minors of Δ which cancel each other. It is, however, easy to avoid such complication by considering simply a term of Δ, any term T, having just t elements in B. Since we know that every $D_kD'_k$ consists of terms of Δ, we see that T will be found once in Δ, t times in $\Sigma D_1D'_1$, $\binom{t}{2}$ times in $\Sigma D_2D'_2$, and so on, and therefore when the signs prefixed to the Σ's are prefixed to T the total

coefficient of T in the left member of (1) will be

$$1-t+\binom{t}{2}-\ .\ .\ .+(-1)^t\binom{t}{t}=0.$$

Thus (1) is established.

It should be observed that this method of proof by giving attention to the individual terms of the determinants present in a relation or an expression is applicable to many apparently complicated situations and resolves them easily. It is employed again in what follows.

3. Let us next state Lecat's extension of (1) to p dimensions. The determinant Δ, of order n, may have any signancy; that is, in the general matrix

$$M=||a_{\alpha_1\alpha_2\ .\ .\ .\ \alpha_p}||_n$$

any desired directions may be taken into account and any left out of account in fixing the sign of a term, so long as an even number are taken into account.[6] And indeed all the relations here dealt with are applicable to permanents as well as to determinants.

In the "leading corner" of M we mark off a "rectangular" space B consisting of the elements common to the first r_1 layers (*tranches*) of the first direction, the first r_2 layers of the second direction, and so on, this rectangular space therefore consisting of all elements $a_{\alpha_1\alpha_2\ .\ .\ .\ \alpha_p}$ or $a(\alpha_1\alpha_2\ .\ .\ .\ \alpha_p)$ in whose locant (*i.e.*, $\alpha_1\alpha_2\ .\ .\ .\ \alpha_p$) the index α_1 has any value from 1 to r_1, the index α_2 has any value from 1 to r_2, and so on, it being provided for convenience of statement that no r shall be less than r_1. In B we can have minors of Δ of all orders from 1 to r_1, and the product of any such minor D_k and its cofactor D'_k consists of terms of Δ. As before, Δ_C will denote a determinant obtained from Δ by replacing all the elements of B by zeros. Lecat's theorem may now be written:

$$\Delta-\Sigma D_1D'_1+\Sigma D_2D'_2-\ .\ .\ .+(-1)^{r_1}\Sigma D_{r_1}D'_{r_1}=\Delta_C. \qquad (2)$$

Without change of language our proof of (1) applies to (2).

[6] The matter of signancy is more fully dealt with in P-way determinants, by L. H. Rice, Amer. Journ. of Math., Vol. XL. (1918), pp. 242–261, esp. at p. 246.

It is noted by Lecat that a sufficient though not a necessary condition for

$$\Delta_C = 0$$

is that

$$\Sigma r_i > n(p-1).$$

We may formulate a proof as follows, first writing the inequality in the form $\Sigma(n-r_i) < n$. Δ_C will vanish if there is no term of Δ without an element in B. If we start to form such a term, we have freedom in the first direction for $n-r_1$ of its elements, freedom in the second direction for $n-r_2$ more elements, and so on, — freedom altogether for $\Sigma(n-r_i)$ elements — and there are n elements to be accommodated.

4. Muir now groups those products $D_k D'_k$ in which the D_k's occupy the same selection of k rows of B and replaces them by a single determinant formed from Δ by replacing the portions of these k rows that project outside B by zeros, a determinant of which obviously this group of products is simply Laplace's expansion, thus converting (1) to a form which we may write (essentially following Muir):

$$\Delta - \Sigma \Delta_1 + \Sigma \Delta_2 - \ldots + (-1)^r \Delta_r = \Delta_C. \tag{3}$$

In the corresponding conversion of (2) made by Lecat, layers of the first direction take the place of rows (which are layers of the first direction in the 2-way case); the portions of these layers projecting beyond B being replaced by zeros in the selected k layers. The result may be written:

$$\Delta - \Sigma \Delta_1 + \Sigma \Delta_2 - \ldots + (-1)^{r_1} \Delta_{r_1} = \Delta_C. \tag{4}$$

5. We come now to axial forms, and first to Muir's theorem that if the main diagonal of a 2-way determinant Δ be taken as the space C, and all the matrix outside it as B, and $\Sigma \Delta_1$ be the sum of the n determinants formed from Δ by replacing *one* main diagonal element by a zero, and so on up to Δ_n, and Δ_C be of course now simply the main diagonal term, then

$$\Delta - \Sigma \Delta_1 + \ldots + (-1)^n \Delta_n = \Delta_C. \tag{5}$$

Cayley's expansion is here depended on for proof.

This relation is extended by Lecat to a p-way determinant Δ in whose matrix an axial space is delimited. It will of course be understood that in a 3-way or cubic determinant two types of axial space exist, for besides that of one dimension consisting of a diagonal line of elements, as

$$a_{111}\ a_{222}\ a_{333}\ .\ .\ .\ a_{nnn},$$

there are axial spaces of two dimensions, as

$$\begin{Vmatrix} a_{111} & a_{122} & a_{13} & . & . & . & a_{1nn} \\ a_{211} & a_{222} & a_{233} & . & . & . & a_{2nn} \\ . & . & . & . & . & . & . \\ a_{n11} & a_{n22} & a_{n33} & . & . & . & a_{nnn} \end{Vmatrix}.$$

In a matrix of four dimensions there are three types, illustrated by the principal axial spaces

$$\| a_{iiii} \|,\ \| a_{ijjj} \|,\ \| a_{ijkk} \|,$$

the latter being of three dimensions. In general, the principal axial space of k dimensions in a matrix of p dimenions is

$$\| a(i_1 i_2\ .\ .\ .\ i_{k-1} i_k i_{k+1} \ldots i_p) \|.\quad i_k = i_{k+1} = \ .\ .\ .\ = i_p.$$

Let this k-way axial space be denoted by C and the remainder of the matrix by B. The complete symmetry of form now makes it immaterial in which of the p directions we replace portions of layers by zeros; let us say that the layers of the mth direction are thus altered, and with the consequent meaning of $\Sigma \Delta_1$, $\Sigma \Delta_2$, etc., and Δ_C, Lecat's extension is:

$$\Delta - \Sigma \Delta_1 + \Sigma \Delta_2 - .\ .\ .\ + (-1)^n \Delta_n = \Delta_C, \qquad (6)$$

his proof being by means of his laws for the development of a determinant in minors lying in an axial space, given in a former paper.

It is noted by Lecat that if the mth direction precede the kth the part of the axial space that lies in any one of the layers will be an axial space of $k-1$ dimensions with respect to the matrix of that layer, otherwise it will be a *couche* of $k-1$ dimensions.[7]

[7] A couche of one dimension is a file; of two dimensions, a square array consisting of the n^2 elements obtained from the general element by allowing two indices independently to take values from 1 to n, while holding fast the other $p-2$ indices at any set of fixed values; and so on, the extreme being a layer, in which but one index is held fast. In the present case, to hold fast the kth or any following index is to hold fast all but the first $k-1$ indices.

6. A highly important generalization is next made by Lecat; he says that the spaces B and C need have no regular form whatsoever, but any elements of the matrix, distributed through it in any manner, may constitute class B, the rest of the elements constituting class C. An arbitrary direction having been chosen, let us denote by $\Sigma\Delta_{BC'}$ the sum of determinants obtained from Δ by replacing in *one* layer containing elements of B the elements of C by zeros; by $\Sigma\Delta_{BC''}$ the sum obtained by doing the same in *two* layers; and so on; by Δ_C the determinant obtained from Δ by replacing all the elements of B by zeros; and, for uniformity, let us denote by Δ_{BC} the determinant Δ itself. Let us suppose that there are elements of B in r layers. Then Lecat's relation is:

$$\Delta_{BC}-\Sigma\Delta_{BC'}+\ .\ .\ .\ +(-1)^r\Delta_{BC^{(r)}}=\Delta_C. \tag{7}$$

For the special case of two dimensions of course the layers are rows (or columns). Lecat gives examples in two, three, and four dimensions; following is his first example in two dimensions:[8]

$$\begin{vmatrix} .&.&c\\ d&.&.\\ g&h&i \end{vmatrix}=\begin{vmatrix} a&b&c\\ d&e&f\\ g&h&i \end{vmatrix}-\begin{vmatrix} a&b&.\\ d&e&f\\ g&h&i \end{vmatrix}-\begin{vmatrix} a&b&c\\ .&e&f\\ g&h&i \end{vmatrix}+\begin{vmatrix} a&b&.\\ .&e&f\\ g&h&i \end{vmatrix}.$$

7. Since no proof of (7) is given by Lecat, we offer a proof which follows the method used in §2. Any term of Δ in which just t elements belong to B will be found once in $\Delta_{BC}(=\Delta)$, t times in $\Sigma\Delta_{BC'}$, $\binom{t}{2}$ times in $\Sigma\Delta_{BC''}$, and so on, and thus its total coefficient will be

$$1-t+\binom{t}{2}-\ .\ .\ .\ +(-1)^t\binom{t}{t}=0.$$

And any term of Δ in which no elements belong to B will be found only in Δ_{BC} and in Δ_C. And as (7) consists wholly of terms of Δ, it is reduced to $0=0$.

8. Coming now to the matter of introducing arbitrary quantities in place of zeros in the foregoing relations, there is first the

[8] In his second example in two dimensions there is some error in the printing for the total coefficient of a is not zero.

theorem of Muir for the case of two dimensions, namely, that in (5) the zeros that replace the main diagonal elements may themselves be replaced by arbitrary quantities provided that these same quantities be also subtracted from the corresponding main diagonal elements in Δ_C; that is, if the main diagonal elements be denoted by $c_{11}, c_{22}, \ldots, c_{nn}$ and the arbitrary quantities by $c_{11}', c_{22}', \ldots, c_{nn}'$, we now have

$$\Delta_C = (c_{11}-c_{11}')(c_{22}-c_{22}')\ldots(c_{nn}-c_{nn}').$$

Lecat states the rule for (6), zeros and elements of the axial space being affected in like manner.

9. Muir also gave a rule for the introduction of arbitrary quantities in (3). Lecat, however, shows by means of an example that the rule is not correct; but does not himself give any rule for this case nor undertake to find a rule for the generalized case of (7).

We proceed to give such a rule, calling it Rule 1; to give a simple example in two dimensions; to prove the correctness of this example; and then to prove the rule.

Rule 1. In the relation (7) the zeros that replace elements of C may themselves be everywhere replaced by arbitrary quantities, one to each place in the matrix, provided that these same quantities be subtracted from the corresponding elements of Δ_C.

For illustration take the example following (7). Let c' and d' be arbitrary quantities, then we shall have:

$$\begin{vmatrix} . & . & c-c' \\ d-d' & . & . \\ g & h & i \end{vmatrix} = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} - \begin{vmatrix} a & b & c' \\ d & e & f \\ g & h & i \end{vmatrix} - \begin{vmatrix} a & b & c \\ d' & e & f \\ g & h & i \end{vmatrix} + \begin{vmatrix} a & b & c' \\ d' & e & f \\ g & h & i \end{vmatrix}.$$

Proof: this is simply the expansion as a sum of determinants with monomial elements of the determinant with binomial elements

$$\begin{vmatrix} a-a & b-b & c-c' \\ d-d' & e-e & f-f \\ g & h & i \end{vmatrix}.$$

Proof of Rule 1. To facilitate the proof let us denote the elements of B by the carrier letter b with the proper locant (suffix) and those of C by c with the proper locant. Form a determinant Δ' from Δ by subtracting an arbitrary quantity $b'(\alpha_1 \ . \ . \ . \ \alpha_p)$ from each $b(\alpha_1 \ . \ . \ . \ \alpha_p)$, and an arbitrary quantity $c'(\alpha_1 \ . \ . \ . \ \alpha_p)$ from each $c(\alpha_1 \ . \ . \ . \ \alpha_p)$ in the r layers containing b's. Expanded as a sum of determinants with monomial elements, Δ' will yield 2^r terms, of which the first will be Δ, the next r will each have one primed layer, the next $\binom{r}{2}$ will each have two primed layers, and so on, up to a single determinant with r primed layers. Now place

$$b'(\alpha_1 \ . \ . \ . \ \alpha_p) = b(\alpha_1 \ . \ . \ . \ \alpha_p),$$

which makes Δ' identical with Δ_C as altered by Rule 1 and makes the expansion identical (signs included) with the altered left member of (7).

This mode of proof supplies also an independent proof of (7) itself, if we place

$$b'(\alpha_1 \ . \ . \ . \ \alpha_p) = b(\alpha_1 \ . \ . \ . \ \alpha_p), \quad c'(\alpha_1 \ . \ . \ . \ \alpha_p) = 0.$$

10. A second rule is immediately suggested:

Rule 2. In the relation (7) the zeros that replace elements of B in Δ_C may themselves be replaced by arbitrary quantities, provided that these same quantities be subtracted from the corresponding elements of the determinants of the left member wherever in the layers containing these elements zeros were substituted for elements of C in forming (7).

For illustration take again the example following (7). Let a'', b'', e'', f'' be arbitrary quantities, then we shall have

$$\begin{vmatrix} a'' & b'' & c \\ d & e'' & f'' \\ g & h & i \end{vmatrix} = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} - \begin{vmatrix} a-a'' & b-b'' & . \\ d & e & f \\ g & h & i \end{vmatrix} - \begin{vmatrix} a & b & c \\ . & e-e'' & f-f'' \\ g & h & i \end{vmatrix}$$

$$+ \begin{vmatrix} a-a'' & b-b'' & . \\ . & e-e'' & f-f'' \\ g & h & i \end{vmatrix}.$$

This is simply the expansion of

$$\begin{vmatrix} a-a' & b-b' & c-0 \\ d-0 & e-e' & f-f' \\ g & h & i \end{vmatrix}$$

with a' then replaced by $a-a''$, etc.

Rule 2, therefore, is established by leaving $b'(\alpha_1 \ldots \alpha_\rho)$ arbitrary in Δ' and then making the substitution

$$b'(\alpha_1 \ldots \alpha_\rho) = b(\alpha_1 \ldots \alpha_\rho) - b''(\alpha_1 \ldots \alpha_\rho).$$

Finally it is to be observed that Rule 1 and Rule 2 may be applied in succession to the relation (7).

THE DIRICHLET PROBLEM

By Norbert Wiener

In a recent paper in the *Comptes rendus*,[1] Lebesgue points out that the Dirichlet problem divides itself into two parts, the first of which is the determination of a harmonic function corresponding to certain boundary conditions, while the second is the investigation of the behaviour of this function in the neighborhood of the boundary. In a paper appearing the same week[2] the author of the present paper made the same remark independently, and went on to give a precise definition of the sense in which the harmonic function depends on the boundary conditions. He proved, moreover, that his method assigns a definite harmonic function to any continuous boundary condition on any bounded set of points in any number of dimensions. In particular, he proved the theorem that the potential corresponding to boundary values 1 on a given bounded set of points and (in three or more dimensions) 0 at infinity determines in his sense a harmonic function generated by a distribution of charge (represented by a Stieltjes integral) over the bounded set. The charge on no region will be negative. The total amount of the charge will be known as the capacity of the boundary set of points. Thus every bounded set of points, whether a surface or not, and indeed whether measurable or not, will have a definite, finite capacity. Moreover, if C_2 is a set of points containing as a part C_1, the capacity of C_2 will at least equal that of C_1.

To return to Lebesgue: like the author, he points out that the solubility of the Dirichlet problem in the classical sense for a given region depends only on the *im Kleinen* properties of the boundary, and may be reduced to the investigation of what he terms the regularity of the individual points of the boundary. A point O of the boundary B of a set of points D is termed regular

[1] January 21, 1924.

[2] *Certain Notions in Potential Theory, Jour. Math. Phys. M. I. T.*, January, 1924.

if whenever $F(P)$ is a function continuous on B, and $f(P)$ is the corresponding function harmonic over D,

$$\lim_{P \to O} f(P) = F(O).$$

This must hold for *every* function f continuous on B.

The oldest fairly general condition for the regularity of a point on a boundary in three-space is that of Poincaré and Zaremba.[3] This is stated by Lebesgue as follows: *A point O of the boundary of a domain D is regular if it is the vertex of a closed conical surface the interior of which is exterior to D.* A result obtained by G. E. Raynor[4] may be stated as follows: *A point O of the boundary of a domain D is regular if there is a number λ greater than zero such that there exist an infinity of spheres S with O as center and such that the measure of the set of points on S and exterior to D exceeds λ times the measure of the surface of S.* This anticipates and contains condition A of Lebesgue, which reads: *The point O is regular if on each sphere with center O and radius r there is a point exterior to D and distant from D by a quantity at least equal to Kr, K being a given positive quantity.* Lebesgue also quotes a condition due to Bouligand,[5] which reads: *The point O is regular if it is the vertex of a conical surface, open or closed, exterior to D.*

All the conditions so far stated, with the exception of that of Raynor, are special cases of one published by H. B. Phillips and the author of the present paper[6] prior to all these conditions except that of Poincaré and Zaremba. In its three-dimensional form, this reads as follows: *The point O is regular whenever there are positive numbers a and b such that if $r<a$ and a sphere S of radius r is constructed with O as center, the content of the projection of the points in S and not in D on some plane exceeds br^2.* In the proof of this theorem, only a denumerable set of spheres S are actually used. If the theorem be stated in the somewhat generalized

[3] Poincaré, *Sur les équations aux derivées partielles de la physique mathématique, Am. J. of M.*, V. 12, 1890; Zaremba, *Sur le principe du minimum, Bull. de l'Ac. des Sc. de Cracovie*, 1909.

[4] *Dirichlet's Problem, Annals of Mathematics*, dated March, 1922, but not actually published until the end of March, 1923.

[5] *Sur le problème de Dirichlet harmonique, Comptes rendus*, January 2, 1924

[6] *Nets and the Dirichlet Problem, Jour. Math. Phys. M. I. T.*, March, 1923.

form involving only a denumerable set of spheres, it will include that of Raynor.

In his paper on *Certain Notions in Potential Theory*, the author gives the following sufficient condition for the regularity of a point: *The point O of the boundary of a domain D is regular if there is a sequence of values of r tending to 0 such that either:*

(*a*) *The inner Lebesgue measure of the points on a sphere of radius r exterior to D in terms of the surface of the sphere itself; exceeds a positive quantity independent of r; or*

(*b*) *The capacity of the part of the boundary of D interior to a sphere of radius r exceeds M times the capacity of the sphere, M being independent of r.*

This condition includes as special cases all those previously mentioned. It is to be noted that it introduces the concept of capacity. This is rendered possible by an antecedent proof that every bounded set of points has a capacity.

Up to this point the theory has dealt with conical points and their immediate generalizations. We next turn to the discussion of cuspidal points and their generalizations. The published work on this subject, so far as is known by the author, is entirely due to Lebesgue,[7] who gave the first example[8] of a simply connected boundary in three-space with an irregular point. Lebesgue's results in this direction have been collected in his recent paper, and read as follows: *If there is a segment of a straight line (or analytic curve) terminating at the point O of the boundary of the domain D, if the segment is entirely exterior to D, and if there exist positive numbers A and B such that if P is a point of the segment, its distance from D exceeds $A\,\overline{OP}^B$, O is regular. If there exist positive numbers A and B such that when P lies on the segment, $\overline{OP} = \overline{OQ}$, and*

$$\overline{PQ} > Ae^{\frac{-B}{\overline{OP}}},$$

Q lies in D, O is irregular. Lebesgue states that he has been unable to generalize these results as he has generalized that of Zaremba, by constructing an infinite set of spheres about the

[7] Mr. Gleason of Princeton, however, verbally communicated to me last year several very interesting results in this direction which he obtained several years ago, before Lebesgue had published similar results.

[8] *Comptes rendus des séances de la société mathématique de France*, 1913.

point O, marking on each the region excluded by D, and rotating each independently about O. He also states no results on cuspidal points reducing to flat tongues.

We now come to conditions of a much more general character. It was stated by Poincaré and repeated by Lebesgue that if a point is regular for a given boundary, it is regular for a boundary entirely interior to the first in the neighborhood of this point. Lebesgue gives the following necessary and sufficient condition for the regularity of the point O: *For a point O of the boundary of a domain D to be regular, it is necessary and sufficient that there exist a function $F(x, y, z)$ continuous at O, attaining its lower bound at O and O only, and such that everywhere in D*

$$\frac{\partial^2 F}{\partial x^2}+\frac{\partial^2 F}{\partial y^2}+\frac{\partial^2 F}{\partial z^2}\leq 0.$$

Other forms of this condition are stated by Lebesgue and Kellogg.[9] They all suffer from the defect of involving the geometrical character of the boundary only in a very indirect and devious manner. From the geometrical point of view, indeed, they are scarcely more than restatements of the regularity of O.

It is the purpose of this paper to develop a complete necessary and sufficient characterization of regular points which shall be at least quasi-geometrical. Its statement reads as follows: *Let O be a point of C, the boundary of an open set of points D. Let λ be any positive quantity less than* 1. *Let γ_n be the capacity of the set of all points Q not belonging to D such that*

$$\lambda^n \leq \overline{OQ} \leq \lambda^{n-1}.$$

Then O is regular or irregular according as

$$\frac{\gamma_1}{\lambda}+\frac{\gamma_2}{\lambda^2}+\ \cdots\ +\frac{\gamma_n}{\lambda^n}+\ \cdots$$

diverges or converges.

In order to prove this theorem, it will be necessary to repeat from the author's previous paper a few theorems relating to the determination of a harmonic function by continuous boundary conditions on an arbitrary boundary. He there proved that if

[9] *An Example in Potential Theory, Proc. Am. Ac. Arts and Sciences*, 1923.

C is any closed bounded set of points, and $F(P)$ any function continuous on this set of points, there is a function $f(P)$ harmonic except on C, and determined by F in the following manner. We first construct[10] a continuous function $g(P)$ defined throughout space and reducing to $F(P)$ on C. We then construct a sequence of boundaries for which the Dirichlet problem is soluble, each within its predecessor, and together sharing no point not on C. This is always possible. We finally form the sequence of harmonic functions assuming the boundary values $g(P)$ on these boundaries. Outside C this sequence will converge to a harmonic function independent of $g(P)$ and the sequence of boundaries, and entirely determined by C and $F(P)$. In the particular case where $F(P)$ is identically 1, $g(P)$ may also be taken as identically 1.

It may be remarked at once that if $f(P)$ and $g(P)$ are any two harmonic functions corresponding to boundary values $F(P)$ and $G(P)$, respectively, on the same boundary C, and if throughout C we have $F(P) \geq G(P)$, then we shall always have $f(P) \geq g(P)$. This may be considerably generalized: let Q be a point interior to the boundary C, and let C' be a boundary containing Q in its interior. Let $F(P) \geq G(P)$ at every point of C in or on C', and let $f(P) \geq g(P)$ at every point of C' interior to C. Then $f(Q) \geq g(Q)$.

If $f(P)$ is the harmonic function corresponding to 1 on the boundary C, if C' is another boundary such that every point of C is on or exterior to C', and if $g(P)$ is the harmonic function corresponding to 1 on C', while both f and g correspond to a potential 0 at infinity, in case the region over which they are defined is unbounded, then $g(P) \geq f(P)$ at any point where both are defined. From this it may be concluded at once that if a given set of points completely surrounds another set, the former set has a capacity at least as large as the latter.

The lemmas of the last two paragraphs are readily proved from the definition given by the author of the sense in which the generalized Dirichlet problem is soluble. They will be presupposed without explicit reference in all that follows.

[10] For the possibility of this construction, see Lebesgue, *Sur le problème de Dirichlet*, *Rendiconti di Palermo*, v. 24 (1907), pp. 371–402.

We now proceed to the proof of our main theorem. We first wish to show that if

$$\frac{\gamma_1}{\lambda}+\frac{\gamma_2}{\lambda^2}+\ \cdots\ +\frac{\gamma_n}{\lambda^n}+\ \cdots$$

diverges, O is regular. It is to be noted that under this condition either

$$\frac{\gamma_1}{\lambda}+\frac{\gamma_3}{\lambda^3}+\ \cdots\ +\frac{\gamma_{2n+1}}{\lambda^{2n+1}}+\ \cdots$$

or

$$\frac{\gamma_2}{\lambda^2}+\frac{\gamma_4}{\lambda^4}+\ \cdots\ +\frac{\gamma_{2n}}{\lambda^{2n}}+\ \cdots$$

diverges.

Let C_n be the set of all points exterior to D in the zone between a sphere of radius λ^n about O as center and a sphere of radius λ^{n-1}. Its capacity, as we have seen, will be γ_n. Now consider the set of points consisting of all points contained in a sphere of radius $\lambda^{n-3/2}$ about O as center and exterior to C_n. We wish to consider the harmonic function corresponding to boundary values O over the sphere and 1 over C_n. This function we shall term $F_n(P)$.

To begin with let C_n consist entirely of regular points. Let the harmonic function $V_n(P)=1$ on C_n and 0 at infinity be represented by the Stieltjes integral

$$\int\int\int \overline{PQ}^{-1}dM(Q).$$

$dM(Q)$ will here be everywhere positive, and will represent the charge over the rectangular parallelepiped $dx\ dy\ dz$. It will vanish when this parallelepiped contains no point of C_n.

The Green's function of the sphere of radius C about O is

$$G_C(P,Q)=\frac{1}{\overline{PQ}}-\sqrt{\frac{1}{C^2+\dfrac{\overline{OP}^2\overline{OQ}^2}{C^2}-2\overline{OP}\ \overline{OQ}\cos(OP,OQ)}}.$$

Let it be noted that

$$G_C(P,Q)\leq 1/\overline{PQ},$$

and that if P and Q both lie at a distance from O not greater than μC $(\mu<1)$,

$$G_C(P,Q)\geq(1-\mu^2)^2/8\overline{PQ}.$$

Hence over a sphere of radius λ^{n-1} about O,

$$\iiint \overline{PQ}^{-1} dM(Q) \geq \iiint G_{\lambda^{n-3/2}}(P, Q) dM(Q)$$
$$\geq \frac{(1-\lambda)^2}{8} \iiint \overline{PQ}^{-1} dM(Q).$$

The function

$$\iiint G_{\lambda^{n-3/2}}(P, Q) dM(Q)$$

is harmonic within our sphere of radius $\lambda^{n-3/2}$ and exterior to C_n It may be verified at once that it vanishes on the sphere. It certainly does not exceed 1 on C_n. Hence

$$F_n(P) \geq \iiint G_{\lambda^{n-3/2}}(P, Q) dM(Q).$$

This can be generalized at once to the case where C_n contains irregular points, by taking a set of regular points C_n' surrounding C_n, assigning boundary values 1 on C_n' and 0 on the sphere, and letting C_n' shrink to C_n.

On a sphere of radius $\lambda^{n+1/2}$ with O as center we shall have

$$F_n(P) \geq \frac{(1-\lambda)^2}{8} \iiint \overline{PQ}^{-1} dM(Q)$$
$$\geq \frac{(1-\lambda)^2}{8} \iiint \tfrac{1}{2} \lambda^{1-n} dM(Q).$$
$$= \frac{(1-\lambda)^2}{16\lambda} \frac{\gamma_n}{\lambda^n}.$$

Now let us consider the potential on a sphere of radius $\lambda^{n+5/2}$ corresponding to a potential of 0 on a sphere of radius $\lambda^{n-3/2}$ about O and a potential 1 on C_n and C_{n+2}. This will be a potential corresponding to a potential at least

$$\frac{(1-\lambda)^2}{16\lambda} \frac{\gamma_n}{\lambda^n}$$

on a sphere of radius $\lambda^{n+1/2}$ about O and 1 on C_{n+2}. Hence it will be at least the sum of

$$\frac{(1-\lambda)^2}{16\lambda} \frac{\gamma_n}{\lambda^n}$$

and the potential due to a boundary potential of 0 on a sphere of radius $\lambda^{n+\frac{1}{2}}$ about O and

$$1-\frac{(1-\lambda)^2}{16\lambda}\frac{\gamma_n}{\lambda^n}$$

on C_{n+2}. In other words, it will at least equal

$$\frac{(1-\lambda)^2}{16\lambda}\frac{\gamma_n}{\lambda^n}+\frac{(1-\lambda)^2}{16\lambda}\frac{\gamma_{n+2}}{\lambda^{n+2}}\left[1-\frac{(1-\lambda)^2}{16\lambda}\frac{\gamma_n}{\lambda^n}\right]$$

$$=1-\left[1-\frac{(1-\lambda)^2}{16\lambda}\frac{\gamma_n}{\lambda^n}\right]\left[1-\frac{(1-\lambda)^2}{16\lambda}\frac{\gamma_{n+2}}{\lambda^{n+2}}\right]$$

on C_{n+4}. In a similar manner, the potential within some sphere about O corresponding to a potential 1 on C_n, C_{n+2}, ... , C_{n+2m} is at least

$$1-\left[1-\frac{(1-\lambda)^2}{16\lambda}\frac{\gamma_n}{\lambda^n}\right]\left[1-\frac{(1-\lambda)^2}{16\lambda}\frac{\gamma_{n+2}}{\lambda^{n+2}}\right]\cdots\left[1-\frac{(1-\lambda)^2}{16\lambda}\frac{\gamma_{n+2m}}{\lambda^{n+2m}}\right]$$

Thus the potential corresponding to a potential 1 on C_n, C_{n+2}, ... , C_{n+2m}, ... assumes continuously the boundary value 1 at O if

$$1-\prod_{m=0}^{\infty}\left[1-\frac{(1-\lambda)^2\gamma_{n+2m}}{16\lambda}\frac{}{\lambda^{n+2m}}\right]=1$$

By a familiar theorem in analysis, this will be the case when and only when

$$\sum_{m=0}^{\infty}\frac{\gamma_{n+2m}}{\lambda^{n+2m}}$$

diverges.

We are now in a position to proceed to a demonstration of the regularity of O. It may be shown immediately by the alternating process, as in my previous paper, that the harmonic function corresponding to boundary values 1 on C_n, C_{n+2}, . . . and 0 on the part of C at a distance at least λ^{n-2} from O is continuous at O. It is easy to show, however, that this latter function is dominated by the harmonic function corresponding to any continuous boundary conditions on C that are non-negative and are 1 over that part of C at a distance from O not exceeding λ^{n-2}.

It follows that any harmonic function corresponding to such boundary conditions on C is continuous at O. However, any positive boundary conditions continuous at O and there alone attaining their maximum value 1 can be reduced to boundary conditions of this type by a modification not exceeding the arbitrarily small quantity ϵ. Thus the harmonic function corresponding to these new boundary conditions cannot have an oscillation of more than ϵ at O, and so must assume its boundary values continuously. Since any continuous boundary condition may be represented linearly in terms of two boundary conditions of this sort, O is regular. Hence O is regular if either

$$\frac{\gamma_1}{\lambda}+\frac{\gamma_3}{\lambda^3}+\ \cdots\ +\frac{\gamma_{2n+1}}{\lambda^{2n+1}}+\ \cdots$$

or

$$\frac{\gamma_2}{\lambda^2}+\frac{\gamma_4}{\lambda^4}+\ \cdots\ +\frac{\gamma_{2n}}{\lambda^{2n}}+\ \cdots$$

diverges, and consequently if

$$\frac{\gamma_1}{\lambda}+\frac{\gamma_2}{\lambda^2}+\ \cdots\ +\frac{\gamma_n}{\lambda^n}+\ \cdots$$

diverges.

We now proceed to the converse of this theorem. For this we need a lemma, to the effect that if O is a regular point of the boundary of the domain D, then given any positive numbers η and R, there is a positive number r less than R such that the potential at O corresponding to a boundary potential 0 at infinity and 1 on the set of points exterior to D and lying at a distance from O between r and R inclusive exceeds $1-\eta$. In the first place, the harmonic function $F(P)$ corresponding to boundary values 0 at infinity and 1 on the part of C, the boundary of D, at a distance from O not exceeding R assumes its boundary value tinuously at O, as may be seen by a comparison of this harmonic function with one corresponding to boundary values continuous on C, never exceeding 1, 1 at O, and 0 on the part of C at a distance at least R from O. Hence there is a sphere about O such that in the part of this sphere within D, $F(P)$ exceeds $1-\eta/2$. Furthermore, there is a smaller sphere about O such that the potential on the larger sphere corresponding to a potential

0 at infinity and 1 on the smaller sphere nowhere exceeds $\eta/2$. Let the radius of the smaller sphere be r. Then the potential on the larger sphere corresponding to a boundary potential of 0 at infinity and 1 on the set of points exterior to D at a distance from O between r and R, inclusive, exceeds $1-\eta$. From this fact, and the fact that it corresponds to boundary values 1 over the set of points within the larger sphere, outside the sphere of radius r, and exterior to D, it may readily be concluded that it exceeds $1-\eta$ at O.

Now suppose that

$$\frac{\gamma_1}{\lambda}+\frac{\gamma_2}{\lambda^2}+\ \cdots\ +\frac{\gamma_n}{\lambda^n}+\ \cdots$$

converges. Then given any number ϵ, it is possible to find an n such that

$$\frac{\gamma_n}{\lambda^n}+\frac{\gamma_{n+1}}{\lambda^{n+1}}+\ \cdots\ +\frac{\gamma_{n+p}}{\lambda^{n+p}}<\epsilon$$

whatever p may be. The potential at O due to a potential 1 on C_m cannot exceed γ_m/λ^m, since no part of the charge on C_m corresponding to this potential is nearer to O than λ^m. Hence the potential at O corresponding to a potential 1 on $C_n, \ldots, C_{n+p}$ cannot exceed

$$\frac{\gamma_n}{\lambda^n}+\ \cdots\ +\frac{\gamma_{n+p}}{\lambda^{n+p}}<\epsilon$$

Since however ϵ is arbitrarily small, the lemma of the last paragraph shows that O is irregular.

We thus have obtained a necessary and sufficient condition for the regularity of O that explicitly involves the boundary. From this all the more special conditions for regularity may be deduced at once. I shall not, however, discuss them all in detail, but shall proceed at once to the theory of monotone cuspidal points of revolution. We shall say that O is a monotone cuspidal point of revolution of the boundary of D if this boundary in the neighborhood of O is a reëntrant surface with equation

$$\phi=f(\rho)$$

in some scheme of spherical coördinates, where $f(\rho)$ is a monotone

increasing function of ρ never exceeding π and vanishing for $\rho=0$. Let us write ϕ_m for the value of ϕ corresponding to $\rho=\lambda^m$, and let us define C_m and γ_m as before. Then C_m will be entirely included in a right circular cylinder with base of radius $\lambda^{m-1}\sin\phi_{m-1}$ and altitude λ^{m-1}, and hence in a prolate spheroid with semi-axes $2\lambda^{m-1}\sin\phi_{m-1}$, $2\lambda^{m-1}\sin\phi_{m-1}$, and $2\lambda^{m-1}$. On the other hand, C will ultimately contain a right circular cylinder with base of radius $\lambda^m\sin\phi_m$ and altitude $\frac{1-\lambda}{2}\lambda^{m-1}$, and hence a prolate spheroid of semi-axes $\frac{\lambda(1-\lambda)}{4}\lambda^{m-1}\sin\phi_m$, $\frac{\lambda(1-\lambda)}{4}\lambda^{m-1}\sin\phi_m$, and $\frac{\lambda(1-\lambda)}{4}\lambda^{m-1}$. The capacities of these spheroids are respectively[11]

$$\frac{4\lambda^{m-1}\cos\phi_{m-1}}{\log\cot\phi_{m-1}/2}$$

and

$$\frac{\lambda^m(1-\lambda)\cos\phi_m}{2\log\cot\phi_m/2}.$$

Hence

$$\frac{4\cos\phi_{m-1}}{\lambda\log\cot\phi_{m-1}/2}\geq\frac{\gamma_m}{\lambda^m}\geq\frac{(1-\lambda)\cos\phi_m}{2\log\cot\phi_m/2}.$$

It then follows from our general theorem that *O is regular or irregular according as*

$$\sum_1^\infty\frac{\cos\phi_m}{\log\cot\phi_m/2}$$

diverges or converges. This series, however, converges or diverges with

$$\sum_1^\infty\frac{-1}{\log\phi_m}$$

since the ratio between corresponding terms tends to the definite limit 1. In particular, if $f(\rho)>A\rho^n(A>0)$, this series dominates a harmonic progression and diverges. On the other hand, if

[11] Jeans, *Electricity and Magnetism*, p. 248.

$f(\rho) < Ae^{-\frac{n}{\rho}}$, this series is dominated by a geometric progression with ratio between 0 and 1, and converges. From these facts, Lebesgue's results as to cuspidal points with finite and exponential order of contact follow at once.

We can generalize this result considerably with the aid of a few lemmas. To begin with, if we replace the boundary in the neighborhood of the cuspidal point which we have just discussed by a certain sequence of zones of one base, the regularity of O will remain unchanged. These zones are chosen in the following manner: the first consists in those points exterior to D at a distance ρ_1 from O. The second consists in those points exterior to D at a distance $\rho_1 - f(\rho_1)$ from O. The third consists in those points at a distance $\rho_1 - f(\rho_1) - f(\rho_1 - f(\rho_1))$ from O, and so on. Let the set of all points exterior to D and at a distance between λ^m and λ^{m-1}, inclusive, from O be C_m as before. Let the set of all points on the zones just mentioned between a distance of λ^m and λ^{m-1} from O, inclusive be C_m'. It will be noted that to each point of C_m, there can be assigned a zone of C_m' or C_{m-1}' bounded by a circle of radius say a, such that the point in question is remote from the furthest point of the zone by less than $3a$. The capacity of the zone will exceed or equal $2a/\pi$. Hence to a potential of 1 on C_m' and C_{m-1}' there will correspond a potential at least $2/3\pi$ on C_m. It may be concluded from this at once by a comparison of the potentials at distant points corresponding respectively to a unit potential on C_m and to a unit potential on C_m' and C_{m-1}' that the capacity of C_m is not greater than $3\pi/2$ times the sum of the capacities of C_m' and C_{m-1}'. From this the regularity of O with respect to the boundary consisting of the zones just defined follows at once.

The next lemma reads as follows: Let $S_1, S_2, \ldots$ be a set of sets of points with capacities $\sigma_1, \sigma_2, \ldots$. Let $T_1, T_2, \ldots$ be a set of sets of points with capacities $\tau_1, \tau_2, \ldots$. Let there be positive numbers a and b such that for every n, $\sigma_n \geq a\,\tau_n$, while for every m and n, the lower bound of the distance between a point of S_m and a point of S_n exceeds b times the upper bound of the distance of a point of T_m from a point of T_n. Then the capacity of the set of points

$$S_1 + S_2 + \ldots$$

is at least as large as the capacity of the set of points

$$T_1+T_2+ \ . \ . \ .$$

multiplied by

$$\frac{ab}{a+b} .$$

The proof of this is simple. Suppose $T_1+T_2+ \ . \ . \ .$ brought to potential 1 by a distribution of charge, and let the charge then found on T_m be ω_m. Transfer this charge to S_m, and it will raise the latter set of points to a potential not exceeding ω_m/a. The remainder of the charge is similarly transferred, and cannot raise any point of S_m to a potential of more than ω_m/b. We thus have a positive distribution of charge over $S_1+S_2+ \ . \ . \ .$ which nowhere on the boundary of this set gives rise to a potential greater than $\Sigma\omega_m(1/a+1/b)$, while the total amount of charge is $\Sigma\omega_m$. A comparison with the charge necessary to bring the whole of $S_1+S_2+ \ . \ . \ .$ to a potential $1/a+1/b$ completes the proof of the theorem.

Combining this theorem with the theorem we have just proved concerning a boundary consisting of spherical caps, and remembering that of all sets of points on the surface of a sphere which have a given area, the spherical cap has the least capacity, we obtain the following theorem: *Let O be a point of the boundary of a region D. Let O' be a monotone cuspidal point of revolution on the boundary of a region D'. Let O' be regular. Let us construct spheres about O and O' respectively as centers with radius r. Let the areas of the parts of D and D' respectively on the surface of these spheres be A and A'. Then if for all sufficiently small values of r we have $A<A'$, O is regular.*

This is the theorem generalizing Lebesgue's result concerning cuspidal points as he generalized Zaremba's result concerning conical points. There is another generalization of the Lebesgue result which resembles rather that of Bouligand concerning flat conical points. Let us return to our sequence of zones of one base terminating in a regular point O. It will be noted that the solid angle subtended by one of these zones will approach 0 as we tend to O, and that consequently these zones will tend in shape to flat discs. Their capacity will hence tend to the capacity of a flat

disc bounded by the same circle, and the regularity of O will not be affected if we replace each zone by the corresponding flat disc. If we replace each disc by a disc in the same plane and with the same center, but with a third the radius, we shall have reduced the capacity of each disc to just a third its original value. If we now rotate each disc about its center in such a manner as to bring them all into a single plane passing through O, the least distance between two points of two adjacent discs will be at least a ninth of the greatest distance between two points of the corresponding zones in the original figure. Hence O will still be regular.

It is manifestly true, as Lebesgue has pointed out, that if two boundaries have a point in common, and one boundary is in the neighborhood of this point entirely exterior to another, then if the point is regular for the exterior boundary, it is also regular for the interior boundary. This indeed follows directly from our fundamental theorem. Thus by a comparison on the one hand with the set of discs we have just discussed, and on the other with a monotone cuspidal point of revolution, we get the following criterion for the regularity of a cuspidal point on a lamina: *Let the boundary of the region D in the neighborhood of a point O consist in a plane lamina containing all the points such that*

$$\theta = f(\rho)$$

and no others, in a scheme of polar coördinates in the plain of the lamina with O as origin. Let $f(\rho)$ be a monotone increasing function of ρ never exceeding π and vanishing for $\rho = 0$. Then O is regular or irregular according as

$$\sum_1^\infty \frac{1}{\log f(\lambda^m)}$$

diverges or converges.

We may indeed obtain a still more general theorem concerning flat cuspidal points. To begin with, a unit charge uniformly distributed over any region with a plane projection of area A with respect to its projection can never produce a greater potential at any point than if the region is a circle of area A, and the point where the potential is taken is at the center of the circle. The

potential at the center of the circle will then be $2\sqrt{\frac{\pi}{A}}$. This means that the capacity of the region with projection of area A is at least $\frac{1}{2}\sqrt{\frac{A}{\pi}}$, since a charge of this amount will correspond to a potential nowhere exceeding 1. Now the capacity of a circular disc of area A is $\frac{2}{\pi}\sqrt{\frac{A}{\pi}}$. Thus the capacity of any region of projection A is at least $\pi/4$ times as great as the capacity of a circular disc of the same area. Combining this fact with the lemma we have just proved concerning a sequence of discs, we get the following theorem: *Let O be a point of the boundary of a region D. Let O' be a monotone cuspidal point of revolution on the boundary of a region D'. Let O' be regular. Let C be the part of the boundary of D lying at a distance of from r to r', inclusive, from O, and let C' be the part of the boundary of D' similarly situated with reference to O'. Then if for all sufficiently small values of r and r', some projection of C exceeds C' in area, O is regular.*

So much for the three-dimensional case: the theory for the n-dimensional case is closely similar. For $n>3$, the fundamental theorem reads as follows: *Let O be a point of C, the boundary of an open set of points D, in a space of n dimensions. Let λ be any positive quantity less than 1. Let γ_m be the capacity of the set of all points Q not belonging to D, such that*

$$\lambda^m \leq \overline{OQ} \leq \lambda^{m-1}.$$

Then O is regular or irregular according as

$$\frac{\gamma_1}{\lambda^{n-2}} + \frac{\gamma_2}{\lambda^{2(n-2)}} + \cdots + \frac{\gamma_m}{\lambda^{m(n-2)}} + \cdots$$

diverges or converges.

Here the proof differs in no essential point from that in the three-dimensional case. In the two-dimensional case, a slight complication is introduced by the fact that the potential due to an isolated charge has a logarithmic singularity at infinity instead of vanishing there. This necessitates a recasting of the definition of capacity. The fundamental existence theorem in my previous

paper is stated in an incorrect fashion, although the subsequent work of the paper is perfectly correct, and follows immediately from the correct theorem. To begin with, let R be a bounded set of points in the plane, and let it contain a point O in its interior. Form the harmonic function $f(P)$ corresponding to the boundary values $\log \overline{OP}$ for points P on R and finite at infinity. Let this function assume the value $-a$ at infinity. Then we shall say that the function

$$\frac{f(P)-\log \overline{OP}+a}{a}$$

corresponds to boundary values 1 on R and behaves like $\log \overline{OP}$ at infinity. This function will always exist if R contains points properly in its interior and is of sufficiently small linear dimensions. The condition that R should contain points properly in its interior is inessential, if we follow out the spirit of my last paper by regarding the harmonic function corresponding to boundary values 1 on R and behaving like $\log \overline{OP}$ at infinity as the lower bound (which will always exist if R is small enough) of the harmonic functions corresponding to boundary values 1 on contours near to R and containing R and behaving like $\log \overline{OP}$ at infinity. The harmonic function thus obtained may readily be proved unique, and independent of O.

The proper form of the existence theorem for capacity in the two-dimensional case is the legitimate conclusion of the argument used in deriving the erroneous theorem of my previous paper, and reads as follows: *Let R be any bounded set of points in the plane, entirely within a rectangle with opposite corners (a_1, a_2) and (b_1, b_2), and let*

$$|a_1-b_1|<1, \quad |a_2-b_2|<1.$$

Let $u(P)$ behave like $\log \overline{OP}$ at infinity, and correspond to boundary values 1 over R. Then there will be a function $M(P)$ representing the charge on that portion of R consisting of points with both coördinates less than those of P. $M(P)$ will be an increasing function of the coördinates of P. Over every point Q exterior to R we shall have

$$u(Q)=\int_{a_1}^{b_1}\int_{a_2}^{b_2}(-\log \overline{PQ})dM(P)$$

Moreover, if the rectangle with opposite vertices (x_1, x_2) *and* (y_1, y_2) *is entirely exterior to* R, *then*

$$\int_{x_1}^{y_1}\int_{x_2}^{y_2} dM(P) = 0,$$

We shall term

$$\int_{a_1}^{b_1}\int_{a_2}^{b_2} dM(P)$$

the capacity of R.

With this definition, our cardinal theorem is the following: *Let* O *be a point of* C, *the boundary of an open set of points* D *in the plane. Let* λ *be any positive quantity less than* 1. *Let* γ_m *be the capacity of the set of all points* Q *not belonging to* D, *such that*

$$\lambda^{2^m} \leq \overline{OQ} \leq \lambda^{2^{m-1}}.$$

Then O *is regular or irregular according as*

$$\gamma_1 + 2\gamma_2 + \ldots + 2^m\gamma_m + \ldots$$

diverges or converges.

The proof follows that in the three-dimensional case step by step. It is of course necessary to take account of the fact that the logarithm assumes positive as well as negative values, but any difficulty from this source may be avoided by confining our attention to the interior of a circle of radius less than $1/2$ about O. The fact that the bounds of $\overline{OQ}$ for points Q on γ_m are λ^{2^m} and $\lambda^{2^{m-1}}$ is accounted for by the fact that the potential at O due to a unit charge at Q will then lie between the numbers $2^m \log \frac{1}{\lambda}$ and $2^{m-1} \log \frac{1}{\lambda}$, which bear to one another a fixed ratio. There is one stage of the proof which perhaps needs a little comment: that in which we determine the potential on a given circle about the origin corresponding to a potential 0 on a circle of larger radius and 1 on γ_m. We shall take the radius of the outer circle as $\lambda^{2^{m-3/2}}$, and that of the smaller circle as $\lambda^{2^{m+1/2}}$. We desire a theorem correlating the potential due to a potential 1 on γ_m and 0 at infinity with the potential in question. As before, we do this by establishing certain inequalities connecting the Green's function of the outer circle with $-\log \overline{PQ}$, the Green's function of the entire plane.

If we write c for the radius of the outer circle, its Green's function will be

$$G(P,Q)=\tfrac{1}{2}\log\left\{\frac{c^2+\frac{\overline{OP}^2\,\overline{OQ}^2}{c^2}-2\overline{OP}\,\overline{OQ}\cos(\overline{OP},\overline{OQ})}{\overline{OP}^2+\overline{OQ}^2-2\overline{OP}\,\overline{OQ}\cos(\overline{OP},\overline{OQ})}\right\}$$

$$=\tfrac{1}{2}\log\left\{\frac{1+\left(c-\frac{\overline{OP}^2}{c}\right)\left(c-\frac{\overline{OQ}^2}{c}\right)}{\overline{PQ}^2}\right\}.$$

We shall always have $G(P,Q)\leq-\log\overline{PQ}$. If $c=\lambda^{2^{m-3/2}}$, while P and Q lie within a circle about O of radius $\lambda^{2^{m-1}}$, we shall have

$$G(P,Q)\geq\tfrac{1}{2}\log\left\{\left[1-\lambda^{(2-\sqrt{2})2^{m-1}}\right]^2\frac{\lambda^{2^{m-1/2}}}{\overline{PQ}^2}\right\}$$

$$\geq\tfrac{1}{2}\log\frac{\lambda^{2^{m-1/2}}}{4\overline{PQ}^2}$$

$$\geq(\log\lambda)2^{m-3/2}-\log 4-\log\overline{PQ}$$

for all sufficiently large values of m. Furthermore,

$$\begin{aligned}-\log\overline{PQ}&\geq-\log 2-\log\lambda^{2^{m-1}}\\&=-\log 2-\log\lambda^{[2^{m-3/2}]\sqrt{2}}\\&=-\log 2-\sqrt{2}\,\log\lambda^{2^{m-3/2}}\end{aligned}$$

Hence for all sufficiently large values of m,

$$(-\log\lambda)2^{m-3/2}+\log 4\leq\frac{-1}{1.4}\log\overline{PQ}$$

It follows that for all sufficiently large values of m, we shall have

$$G(P,Q)\geq-.2\log\overline{PQ}.$$

With this we are in a position to proceed with our two-dimensional theorem in a fashion exactly paralleling that employed in the demonstration of our theorem in three dimensions.

The classical sufficient condition for the solubility of the plane Dirichlet problem is that of Lebesgue[12]. It may be stated as follows: *A point O of the boundary of a region D is regular unless it is possible to construct a circle of arbitrarily small radius about O*

[12] *Rendiconti di Palermo*, loc. cit.

which will lie entirely interior to D. This is completely contained in two conditions, given by Phillips and the author, and the author alone, respectively, which are entirely parallel to the three-dimensional conditions given earlier in this paper. They may readily be deduced from the general condition here given, the series whose divergence is to be investigated reducing to the form

$$A+A+\ \ldots\ +A+\ \ldots$$

A plane problem of particular interest has been discussed by Kellogg in the memoir already cited. He discusses the boundary consisting of all points on the segment (0, 1) which have a representation in ternary fractions not containing the digit 2, and shows that every point of it is regular. If we wish to bring this under our general theorem, we must first show that this set has a non-zero capacity, or in other words that it is possible to distribute over it a positive charge without producing anywhere an infinite potential. This we do as follows: like Kellogg, we form the function $f(x)$ which has the value 0 for $x=0$, 1 for $x=1$, 1/2 for $1/3 \leq x \leq 2/3$, 1/4 for $1/9 \leq x \leq 2/9$, 3/4 for $7/9 \leq x \leq 8/9$, and so on. Then the function

$$-\int_0^1 \log x\, df(x)$$

is bounded, for it never exceeds

$$\log 6+1/2 \log 18+1/4 \log 54+\ \ldots$$

$$=\log 2(1+1/2+1/4+\ldots)+\log 3\left(1+1+\frac{3}{4}+\frac{4}{8}+\frac{5}{16}+\ldots\right)$$

$$\leq 2 \log 2 + 5 \log 3.$$

Hence *a fortiori* the function

$$-\int_0^1 \log \overline{OP}\, df(O)$$

is bounded.

If the capacity of Kellogg's set of points is at least C, the capacity of each non-null third is at least $\frac{C}{1+C \log 3}$, of each non-null ninth $\frac{C}{1+C \log 9}$, and so on indefinitely. Moreover if O is a point of Kellogg's set, γ_m will contain at least one non-null

$1/3^{2+k}$th for all sufficiently large values of m, where $1/3^{k-1} \geq \lambda^{m-1} \geq 1/3^k$. Hence the test-series dominates

$$+ \ldots + \frac{2^m}{1-2^{m-1}C \log \lambda + \log 27} + \frac{2^{m+1}}{1-2^m C \log \lambda + \log 27} + \ldots$$

which diverges. Thus Kellogg's case comes under our general theory, even though it can be treated somewhat more elegantly by Kellogg's direct methods[13].

[13] Since this paper has gone to the printer, the author has become more closely acquainted with the very important researches of M. Georges Bouligand. These go back to his Paris thesis of 1914 on the Green and Neumann functions of the cylinder, and culminate in a paper in the *Comptes rendus* for March 24, 1924, yielding conditions of regularity comparable in generality to those here given. In the latter paper are contained very many references to his previous work, which has appeared for the most part in the *Comptes rendus*. His treatment depends upon the discussion of infinite regions, and the use of inversions and Kelvin transformations.

ON TAUTOCHRONOUS MOTION

By S. D. ZELDIN

The differential equation

$$F \equiv y''' + A(x, y, y')y''^2 + B(x, y, y')y'' + C(x, y, y') = 0,$$

where A, B and C are arbitrary functions of x, y and y', represents the singly infinite system of tautochrones having the following property: At any point (x, y) of the plane in a given direction y' due to a positional force, each tautochrone of the system corresponds to a different value of the time constant. The focal curve of this system is a bicircular quartic having a flecnode at the point, the stationary tangent at that point having the given direction y'. An additional property of this system is — that it is converted by the general point transformation into a system of the same type.[1]

It is clear, however, that the forms of A, B and C will differ for different point transformations. In this paper we determine the forms of A, B and C for all the groups of point transformations of the plane. The groups that we are here considering have all been determined by S. Lie.[2]

According to Lie's theory, the equation $F=0$ is invariant under a given infinitesimal transformation

$$\xi(x, y)\frac{\partial f}{\partial x} + \eta(x, y)\frac{\partial f}{\partial y},$$

if

$$\delta F \equiv \frac{\partial F}{\partial x}\delta x + \frac{\partial F}{\partial y}\delta y + \frac{\partial F}{\partial y'}\delta y' + \frac{\partial F}{\partial y''}\delta y'' + \frac{\partial F}{\partial y'''}\delta y''' = 0,$$

where

$$\delta x = \xi(x, y)\delta t,\ \delta y = \eta(x, y)\delta t,\ \delta y^{(i)} = \frac{d(\delta y^{(i-1)})}{dx} - y^{(i)}\frac{d(\delta x)}{dx}$$

In what follows we consider the twenty-seven groups of point transformations of the plane. Each group is given by its infinitesimal transformations and its differential invariants.

[1] H. W. Reddick, Am. Journal of Math., Vol. 32, p. 376.

[2] Mathematische Annalen, t. XXXII, p. 213; See also, A. Guldberg, Rend. del Circolo Mat. di Palermo, t. XXI, p. 67.

1. The one-parameter group

$$p\ \left(p=\frac{\partial f}{\partial x}\right),$$

with its differential invariants

$$\phi_1=y,\quad \phi_2=y',$$

has for its infinitesimal increments

$$\delta x=\delta t,\quad \delta y=\delta y'=\delta y''=\delta y'''=0.$$

Therefore, it leaves the equation $F=0$ invariant, if

$$\delta F\equiv \delta x.(A_x y''^2+B_x y''+C_x)=0,$$

where

$$A_x=\frac{\partial A}{\partial x},\quad B_x=\frac{\partial B}{\partial x},\quad C_x=\frac{\partial C}{\partial x}.$$

Equating to zero the coefficients of the different powers of y'', we find that

$$A_x=B_x=C_x=0.$$

In other words, A, B and C must be functions of y and y' alone.

2. The two-parameter group

$$q,\ xp+yq\quad \left(p=\frac{\partial f}{\partial x},\ q=\frac{\partial f}{\partial y}\right),$$

with its two differential invariants

$$\phi_1=y''x,\quad \phi_2=y'''x^2$$

leaves the equation $F=0$ invariant, if each of its transformations leaves it invariant.

The transformation q admits the equation $F=0$, if

$$A_y=B_y=C_y=0. \tag{1}$$

The transformation $xp+yq$ gives

$$\delta x=x\delta t,\quad \delta y=y\delta t,\quad \delta y'=0,\quad \delta y''=-y''\delta t,\quad \delta y'''=-2y'''\delta t.$$

Substituting these values in the equation

$$\delta F\equiv\frac{\partial F}{\partial x}\delta x+\frac{\partial F}{\partial y}\delta y+\frac{\partial F}{\partial y'}\delta y'+\frac{\partial F}{\partial y''}\delta y''+\frac{\partial F}{\partial y'''}\delta y'''=0,$$

we obtain

$$xA_x y''^2+(xB_x+B)y''+xC_x+2C=0,$$

and equating the coefficients of y''^2, y'' and $(y')^0$ to zero, we get the equations

$$A_x=0, \quad xB_x+B=0, \quad xC_x+2C=0. \tag{2}$$

Solving equations (1) and (2) simultaneously, we find that the equation $F=0$ is invariant under the given two-parameter group, if

$$A=A(y'), \quad B=\frac{1}{x}B_1(y'), \quad C=\frac{1}{x^2}C_1(y'),$$

where A, B_1 amd C_1 are arbitrary functions of y' alone.

3. The two-parameter group

$$q, \quad yq,$$

with its two differential invariants

$$\phi_1=x, \quad \phi_2=\frac{y''}{y'},$$

gives for the transformation q

$$\delta x=0, \quad \delta y=\delta t, \quad \delta y'=\delta y''=\delta y'''=0.$$

Therefore,

$$A_y=B_y=C_y=0.$$

Now the transformation yq gives

$$\delta x=0, \quad \delta y=y\delta t, \quad \delta y'=y'\delta t, \quad \delta y''=y''\delta t, \quad \delta y'''=y'''\delta t.$$

Substituting these values in the equation $\delta F=0$, we get

$$A+y'A_{y'}=0, \quad B_{y'}=0, \quad y'C_{y'}-C=0.$$

Therefore, we must have

$$A=\frac{1}{y}A_1(x), \quad B=B_1(x), \quad C=y'C_1(x),$$

where A_1, B_1 and C_1 are arbitrary functions of x alone.

4. The three-parameter group

$$p, \quad q, \quad yq,$$

with its two differential invariants

$$\phi_1=\frac{y''}{y'}, \quad \phi_2=\frac{y'''}{y'}.$$

will leave the equation $F=0$ invariant, if

$$A=\frac{k_1}{y'}, \quad B=k_2, \quad C=k_3y',$$

where k_1, k_2 and k_3 are arbitrary constants. This can be proven by the aid of articles 1 and 3.

5. The three-parameter group

$$p, \quad q, \quad xp+ayq \quad (a\neq 0, 1),$$

with its two differential invariants

$$\phi_1=y''/(y')^{\frac{a-2}{a-1}}, \quad \phi_2=y'''/(y')^{\frac{a-3}{a-1}},$$

will leave the equation $F=0$ invariant, if

$$A=k_1/y', \quad B=k_2/(y')^{\frac{1}{a-1}}, \quad C=k_3/(y')^{\frac{a-3}{a-1}},$$

where k_1, k_2 and k_3 are arbitrary constants. This can be proven as follows:

The transformations p and q leave $F=0$ invariant, if

$$A_x=B_x=C_x=A_y=B_y=C_y=0. \tag{3}$$

Now, the transformation $xp+ayq$ gives

$$\delta x=x\delta t, \quad \delta y=ay\delta t, \quad \delta y'=(a-1)y'\delta t, \quad \delta y''=(a-2)y''\delta t,$$
$$\delta y'''=(a-3)y'''\delta t.$$

Substituting these expressions in the equation $\delta F=0$, we get

$$A+y'A_{y'}=0, \quad B+(a-1)y'B_{y'}=0. \quad (a-1)y'C_{y'}-(a-3)C=0.$$

These last equations together with equations (3) give

$$A=k_1/y', \quad B=k_2/(y')^{\frac{1}{a-1}}, \quad C=k_3/(y')^{\frac{a-3}{a-1}}$$

6. The three-parameter group

$$q, \quad yq, \quad y^2q,$$

with its two differential invariants

$$\phi_1=x, \quad \phi_2=\frac{y'y'''-3/2(y'')^2}{(y')^2}.$$

has the infinitesimal increments (with regard to y^2q alone)

$$\delta x=0,\quad \delta y=y^2\delta t,\quad \delta y'=2yy'\delta t,\quad \delta y''=[2yy''+2(y')^2]\delta t,$$
$$\delta y'''=(2yy'''+6y'y'')\delta t.$$

Substituting these values of the infinitesimal increments in the equation $\delta F=0$, and making use of the results obtained in article 3, we find that $F=0$ is invariant under this group, if

$$A=-3/2y',\quad B=0,\quad C=y'C(x).$$

7. The three-parameter group

$$p+q,\quad xp+yq,\quad x^2p+y^2q,$$

with its two differential invariants

$$\phi_1=(x-y)y''(y')^{-3/2}+2[(y')^{1/2}+(y')^{-1/2}]$$
$$\phi_2=(x-y)^2y'''(y')^{-2}+6\phi_1[(y')^{1/2}+(y')^{-1/2}]-6(y'+y'^{-1}),$$

will leave the equation $F=0$ invariant, if

$$A=0,\quad B=\frac{6(y'+1)+c_2(y')^{1/2}}{x-y},$$
$$C=\frac{6y'(y'^2+1)+2c_2(y')^{3/2}(y'-1)+c_3(y')^2}{(x-y)^2},$$

where c_2 and c_3 are arbitrary constants. In order to prove, this, let us, first, find the forms of A, B and C, when $F=0$ is invariant under $p+q$ alone. Here

$$\delta x=\delta t,\quad \delta y=\delta t,\quad \delta y'=\delta y''=\delta y'''=0.$$

Substituting these values in the equation $\delta F=0$, we find that

$$A_x+A_y=0,\quad B_x+B_y=0,\quad C_x+C_y=0. \qquad (a)$$

Now, the transformation $xp+yq$ gives

$$\delta x=x\delta t,\quad \delta y=y\delta t,\quad \delta y'=0,\quad \delta y''=-y''\delta t,\quad \delta y'''=-2y'''\delta t.$$

Substituting these values in the equation $\delta F=0$, we find that, in order that the equation $F=0$ be invariant under $p+q$ and $xp+yq$, the functions A, B and C must satisfy also the equations

$$\left.\begin{aligned} xA_x+yB_y&=0,\\ xB_x+yB_y+B&=0,\\ xC_x+yC_y+2C&=0.\end{aligned}\right\} \qquad (b)$$

Finally, the transformation x^2p+y^2q has for its infinitesimal increments

$$\delta x = x^2\delta t, \quad \delta y = y^2\delta t, \quad \delta y' = (2yy'-2xy')\delta t,$$
$$\delta y'' = (2y'^2+2yy''-2y'-4xy'')\delta t,$$
$$\delta y''' = (6y'y''+2yy'''-6y''-6xy''')\delta t.$$

Substituting these values in the equation $\delta F=0$, we find that A, B and C must satisfy the equations

$$\left.\begin{aligned} &x^2A_x+y^2A_y+(2yy'-2xy')A_{y'}=0,\\ &4y'(y'-1)A+x^2B_x+y^2B_y+(2yy'-2xy')B_{y'}+6(y'-1)=0,\\ &(2y'^2-2y')B+x^2C_x+y^2C_y+(2yy'-2xy')C_{y'}=0. \end{aligned}\right\} \qquad (c)$$

Solving the equations (a), (b) and (c) simultaneously, we find that $F=0$ is invariant under the given three-parameter group, if

$$A=0, \quad B=\frac{6(y'^2+1)+C_2(y')^{1/2}}{x-y},$$

and

$$C=\frac{6y'(y'^2+1)+2c_2(y')^{3/2}(y'-1)+c_3(y')^2}{(x-y)^2},$$

where c_2 and c_3 are arbitrary constants.

8. The three-parameter group

$$p, \quad xp+yq, \quad x^2p+2xyq,$$

with its two differential invariants

$$\phi_1=2yy''-(y')^2, \quad \phi_2=y^2y''',$$

has the following infinitesimal increments:

for p, $\delta x=\delta t$, $\delta y=\delta y'=\delta y''=\delta y'''=0$,

for $xp+yq$, $\delta x=x\delta t$, $\delta y=y\delta t$, $\delta y'=0$, $\delta y''=-y''\delta t$, $\delta y'''=-2y'''\delta t$,

for $x^2p+2xyq$, $\delta x=x^2\delta t$, $\delta y=2xy\delta t$, $\delta y'=2y\delta t$, $\delta y''=(2y'-2xy')\delta t$, $\delta y'''=-4xy'''\delta t$.

The equation $F=0$ is invariant under p, if

$$A_x=B_x=C_x=0; \qquad (e)$$

it is invariant under $xp+yq$, if

$$\left.\begin{aligned} &xA_x+yA_y=0, \\ &xB_x+yB_y+B=0, \\ &xC_x+yC_y+2C=0; \end{aligned}\right\} \qquad (f)$$

and it is invariant under $x^2p+2xyq$, if

$$\left.\begin{aligned} &x^2A_x+2xyA_y+2yA_{y'}=0 \\ &x^2B_x+2xyB_y+2yB_{y'}+4y'A+2xB=0, \\ &x^2C_x+2xyC_y+2yC_{y'}+2y'B+4xC=0. \end{aligned}\right\} \qquad (j)$$

Solving equations (e), (f) and (j) simulataneously, we find that the equation $F=0$ is invariant under this three-parameter group, if

$$A=k_1, \quad B=\frac{1}{y}[k_2-k_1(y')^2],$$

$$C=\frac{1}{y^2}\left[-\frac{k_1}{4}(y')^2+\frac{k_2}{2}(y')^2+k_3\right]$$

where k_1, k_2 and k_3 are arbitrary constants.

9. The three-parameter group

$$p, \quad q, \quad xp+yq,$$

with its two differential invariants

$$\phi_1=y', \quad \phi_2=\frac{y'''}{(y'')^2},$$

leaves equation $F=0$ invariant, if the following equations are satisfied:

$$\begin{aligned} &A_x=B_x=C_x=A_y=B_y=C_y=0, \\ &xA_x+yA_y=0, \\ &xB_x+yB_y+B=0, \\ &xC_x+yC_y+2C=0.^3 \end{aligned}$$

Therefore, we must have

$$A=A(y'), \quad B=0, \quad C=0.$$

[3] For proof, see articles 1, 2, 8.

10. The four-parameter group

$$p,\quad q,\quad yq,\quad xp,$$

with its two differential invariants

$$\phi_1=\frac{y'y'''}{(y'')^2},\quad \phi_2=\frac{(y')^2\ y^{IV}}{(y'')^3}$$

will leave the equation $F=0$ invariant, if A, B and C satisfy the equations

$$A_x=B_x=C_x=A_y=B_y=C_y=0,$$
$$A+y'A_{y'}=0,\quad B_{y'}=0,\quad y'C_{y'}-C=0,$$
$$B=0,\quad 3C-y'C_{y'}=0.$$

Therefore, we must have

$$A=k_1/y',\quad B=C=0,$$

where k_1 is an arbitrary constant.

11. The four-parameter group

$$p,\quad q,\quad yq,\quad y^2q,$$

with its two differential invariants

$$\phi_1=\frac{y'y'''-3/2(y'')^2}{(y')^2},\qquad \phi_2=\frac{(y')^2y^{IV}-4y'y''y'''+3(y'')^2}{(y')^3},$$

leaves invariant the equation $F=0$, if A, B and C satisfy the equations

$$A_x=A_y=B_x=B_y=C_x=C_y=0,$$
$$A+y'A_{y'}=0,\quad B_{y'}=0,\quad y'C_{y'}-C=0,$$
$$3+y'A=0,\quad B=0.$$

Therefore

$$A=-3/y'.\quad B=0.\quad C=ky'.$$

12. The five-parameter group

$$p,\quad yq,\quad xp,\quad x^2p+xyq,$$

with its two differential invariants

$$\phi_1=\frac{yy'''+3y'y''}{y^{1/2}(y'')^{3/2}},\quad \phi_2=\frac{3y\ y''y^{IV}-4y(y''')^2}{(y'')^3},$$

leaves invariant the equation $F=0$, if A, B and C satisfy the equations

$$A_x=B_x=C_x=0,$$
$$A+y'A_{y'}+yA_y=0,\quad y'B_{y'}+yB_y=0,\quad yC_y+y'C_{y'}-C=0,$$
$$A+y'A_{y'}=0,\quad B-y'B_{y'}=0,\quad 3C-y'C_{y'}=0,$$
$$A_{y'}+xA_y=0,\quad -3+2xB+(y-xy')B_{y'}+xyB=0,$$
$$xyC_y+(y-xy')C_{y'}+5xC=0.$$

Solving the above equations, we find that

$$A=0,\quad B=3y'/y,\quad C=0.$$

13. The five-parameter group

$$p,\quad q,\quad xq,\quad yp,\quad xp-yq,$$

with its two differential invariants

$$\phi_1=\frac{3y''y^{\text{IV}}-5(y''')^2}{(y'')^{8/3}},\quad \phi_2=\frac{3(y'')^2y^{\text{V}}-15y''y'''y^{\text{IV}}+\frac{40}{3}(y''')^3}{(y'')^4},$$

leaves the equation $F=0$ invariant, if A, B and C satisfy the equations

$$A_x=A_y=B_x=B_y=C_x=C_y=0,$$
$$A_{y'}=B_{y'}=C_{y'}=0,$$
$$3+2y'A=0,\quad B=C=A=0.$$

But the equations involving A are *inconsistent*, therefore, the equation $F=0$ is not invariant under the given five-parameter group.

14. The five-parameter group

$$p,\quad q,\quad yq,\quad xp,\quad y^2q,$$

with its two differential invariants

$$\phi_1=\frac{\frac{y^{\text{IV}}}{y'}-4\frac{y''y'''}{(y')^2}+3\frac{(y'')^3}{(y')^3}}{\left[\frac{y'''}{y'}-3/2\left(\frac{y''^2}{y'}\right)\right]^{/2}},$$

$$\phi_2=\frac{\dfrac{y^{V}}{y'}-5\dfrac{y''y^{IV}}{(y')^2}-4\dfrac{(y''')^2}{(y')^2}+17\dfrac{(y'')^2y'''}{(y')^3}-9\left(\dfrac{y''}{y'}\right)^4}{\left[\dfrac{y'''}{y'}-3/2\left(\dfrac{y''}{y'}\right)^2\right]^2},$$

admits the equation $F=0$, if A, B and C satisfy the equations

$$A_x=A_y=B_x=B_y=C_x=C_y=0,$$
$$A+y'A_{y'}=0,\quad B_{y'}=0,\quad y'C_{y'}-C=0,$$
$$3+y'A=0,\quad B=0,\quad 3C-y'C_{y'}=0.$$

Solving these equations, we find that

$$A=-3/y',\quad B=C=0.$$

15. The six-parameter group

$$p,\quad q,\quad xq,\quad yq,\quad xp,\quad yp,$$

with its two differential invariants

$$\phi_1=\frac{3(y'')^2y^{V}-15y''y'''y^{IV}+40/3(y''')^3}{[3y''y^{IV}-5(y''')^2]^{3/2}},$$

$$\phi_2=\frac{3(y'')^3y^{VI}-21(y'')^2y'''y^{V}+35y''(y''')^2y^{IV}-35/3(y''')^4}{[3y''y^{IV}-5(y''')^2]^3}$$

does not admit the equation $F=0$, since its four transformations

$$p,\quad q,\quad xq,\quad yp,$$

give rise to inconsistent equations (see Article 13).

16. The six-parameter group

$$p,\quad q,\quad yq,\quad xp,\quad y^2q,\quad x^2p,$$

with its two differential invariants

$$\phi_1=\frac{4\omega\omega''-5(\omega')^2}{\omega^3},\quad \phi_2=\frac{4\omega^2\omega'''-18\omega\omega'\omega''+15(\omega')^3}{\omega^{9/2}},$$

where

$$\omega=\frac{y'''}{y'}-3/2\left(\frac{y''}{y'}\right)^2,\qquad \omega'=\frac{d\omega}{dx},\quad \omega''=\frac{d^2\omega}{dx^2},\quad \omega'''=\frac{d^3\omega}{dx^3},$$

leaves $F=0$ invariant, if the equations

$$A_x=A_y=B_x=B_y=C_x=C_y=0,$$
$$A+y'A_{y'}=0,\quad B_{y'}=0,\quad y'C_{y'}-C=0,$$
$$3+y'A=0,\quad B=0,\quad 3C-y'C_{y'}=0,$$
$$3+2y'A=0$$

are satisfied. But these equations are inconsistent. Therefore, the equation $F=0$ is not invariant under the given six-parameter group.

17. The eight-parameter group

$$p,\quad q,\quad xq,\quad yq,\quad xp,\quad yp,\quad x^2p+xyq,\quad xyp+y^2q,$$

with its two differential invariants

$$\phi_1=\frac{2\rho_3\rho_5-35\rho_2\rho_3^2-7(\rho_4-5/3\rho_2^2)^2}{\rho_3^{8/3}},$$

$$\phi_2=\frac{\rho_3\left(\rho_6-84\rho_3\rho_4+\frac{245}{3}\rho_2^3\right)-12\left(\rho_5-\frac{35}{2}\rho_2\rho_3\right)\times\left(\rho_4-\frac{5}{3}\rho_2^2\right)+\frac{28}{3}\left(\rho_4-\frac{5}{3}\rho^2_2\right)^2}{\rho_3^3},$$

where

$$\rho_2=3y''y^{IV}-4(y''')^2,\quad \rho_3=\frac{1}{3}[3(y'')^2y^V-45y''y'''y^{IV}+40(y''')^3],$$
$$\rho_4=3(y'')^3y^{VI}-24(y'')^2y'''y^V+60y''(y''')^2y^{IV}-40(y''')^4,$$
$$\rho_5=9(y'')^4y^{VII}-105(y'')^3y'''y^{VI}+420(y'')^2y^V-700y''(y''')^2y^{IV}+\frac{1120}{3}(y''')^5,$$
$$\rho_6=27(y'')^5y^{VIII}-48y'''\rho_5-24.35(y''')^2\rho_4-16.40(y''')^3\rho_3-2800(y''')\rho_2-\frac{8.280}{3}(y''')^6,$$

does not admit the equation $F=0$. (See Article 13.)

18. The r-parameter group

$$X_1 q, \quad X_2 q, \quad \ldots, \quad X_r q \quad (X_i \text{ is a function of } x),$$

with its two differential invariants

$$\phi_1 = x, \quad \phi_2 = \begin{vmatrix} X_1 & X_1' & X_1'' & \ldots & X_1^{(r)} \\ X_2 & X_2' & X_2'' & \ldots & X_2^{(r)} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ X_r & X_r' & X_r'' & \ldots & X_r^{(r)} \\ y & y' & y'' & \ldots & y^{(r)} \end{vmatrix},$$

leaves the equation $F=0$ invariant, if A, B and C satisfy the following systems of equations:

$$\left.\begin{array}{l} X_1 A_y + X_1' A_{y'} = 0, \\ X_2 A_y + X_2' A_{y'} = 0, \\ \ldots\ldots\ldots \\ X_r A_y + X_r' A_{y'} = 0. \end{array}\right\} \text{(I)} \qquad \left.\begin{array}{l} X_1 B_y + X_1' B_{y'} + 2X_1' A = 0, \\ X_2 B_y + X_2' B_{y'} + 2X_2' A = 0, \\ \ldots\ldots\ldots\ldots \\ X_r B_y + X_r' B_{y'} + 2X_r' A = 0. \end{array}\right\} \text{(II)}$$

$$\left.\begin{array}{l} X_1 C_y + X_1' C_{y'} + X_1'' B + X_1''' = 0, \\ X_2 C_y + X_2' C_{y'} + X_2'' B + X_2''' = 0, \\ \ldots\ldots\ldots\ldots\ldots \\ X_r C_y + X_r' C_{y'} + X_r'' B + X_r''' = 0. \end{array}\right\} \text{(III)}$$

19. The $(r+1)$-parameter group

$$yq, \quad X_1 q, \quad X_2 q, \quad \ldots, \quad X_r q,$$

with its two differential invariants

$$\phi_1 = x, \quad \phi_2 = \frac{d \log D}{dx},$$

where

$$D = \begin{vmatrix} X_1 & X_1' & \ldots & X_1^{(r)} \\ X_2 & X_2' & \ldots & X_2^{(r)} \\ \ldots & \ldots & \ldots & \ldots \\ X_r & X_r' & \ldots & X_r^{(r)} \\ y & y' & \ldots & y^{(r)} \end{vmatrix},$$

admits the equation $F=0$, if A, B and C satisfy equations (I), (II), (III), and also the equations

$$A+yA_y+y'A_{y'}=0, \quad yB_y+y'B_{y'}=0, \quad yC_y+y'C_{y'}-C=0.$$

20. The $(r+1)$-parameter group

$$p, \quad X_1q, \quad X_2q, \ \ldots \ , \quad X_rq,$$

with its two differential invariants

$$\phi_1=cy+c_1y'+ \ \ldots \ +c_ry^{(r)}, \quad \phi_2=\frac{d\phi_1}{dx},$$

leaves the equation $F=0$ invariant, if the equation (I), (II). (III), and

$$A_x=B_x=C_x=0$$

are satisfied by A, B and C.

21. The $(r+2)$-parameter group

$$p, \ yq, \ X_1q, \ X_2q, \ \ldots \ , \ X_rq,$$

with its two differential invariants

$$\phi_1=\frac{1}{\phi}\frac{d\phi}{dx}, \qquad \phi_2=\frac{1}{\phi}\frac{d^2\phi}{dx^2},$$

where

$$\phi=c_1y+c_2y'+c_3y''+ \ \ldots \ +c_{r+1}y^{(r)},$$

admits the equation $F=0$, if A, B and C satisfy the equations (I), (II), (III), and the equations

$$A_x=B_x=C_x=0, \qquad A+yA_y+y'A_{y'}=0,$$
$$yB_y+y'B_{y'}=0, \qquad yC_y+y'C_{y'}-C=0.$$

22. The $(r+2)$-parameter group

$$q, \ xq, \ \ldots \ , \ x^{r-1}q, \ p, \ xp+ayq \ (a\neq r),$$

with its two differential invariants

$$\phi_1=\frac{y^{(r+1)}}{(y^{(r)})^{\frac{a-r-1}{a-r}}}, \qquad \phi_2=\frac{y^{(r+2)}}{(y^{(r)})^{\frac{a-r-2}{a-2}}},$$

does not admit the equation $F=0$, for, the equations determining A, B and C are inconsistent.

23. The $(r+2)$-parameter group

$$q,\ xq,\ \ldots,\ x^{r-1}q,\ p,\ xp+ryq,$$

with its two differential invariants

$$\phi_1=y^{(r)},\quad \phi_2=\frac{y^{(r+2)}}{(y^{(r+1)})^2},$$

does not admit the equation $F=0$ for the same reason as the group of Article 24.

24. The $(r+2)$-parameter group

$$q,\ xq,\ \ldots,\ x^{r-1}q,\ p,\ xp+(ry+x^r)q,$$

with its two differential invariants

$$\phi_1=y^{(r+1)}.e^{\frac{y^{(r)}}{r!}},\qquad \phi_2=y^{(r+2)}.e^{2\frac{y^{(r)}}{r!}}$$

does not admit the equation $F=0$, for, the equations derived to determine A, B and C are inconsistent.

25. The $(r+3)$-parameter group

$$q,\ xq,\ \ldots,\ x^{r-1}q,\ yq,\ p,\ xp,$$

with its two differential invariants

$$\phi_1=\frac{y^{(r)}y^{(r+2)}}{[y^{(r+1)}]^2},\qquad \phi^2=\frac{[y^{(r)}]^2y^{(r+3)}}{[y^{(r+1)}]^3}.$$

does not admit the equation $F=0$.

26. The $(r+3)$-parameter group

$$q,\ xq,\ \ldots,\ x^{r-1}q,\ p,\ 2xp+(r-1)yq,\ x^2p+(r-1)xyq,$$

with its two differential invariants

$$\phi_1=[y^{(r)}]^{-\frac{2(r+3)}{r+1}}[(r+1)y^{(r)}y^{(r+2)}-(r+2)(y^{(r+1)})^2],$$

$$\phi_2=[y^{(r)}]^{-\frac{3(r+3)}{r+1}}[(r+1)^2(y^{(r)})^2y^{(r+3)}-3(r+1)(r+3)y^{(r)}y^{(r+1)}y^{(r+2)}+2(r+2)(r+3)(y^{(r+1)})^3]$$

does not leave the equation $F=0$ invariant.

27. The $(r+4)$-parameter group

$$q,\ xq,\ \ldots,\ x^{r-1}q,\ yq,\ p,\ xp,$$
$$x^2p+(r-1)xyq,$$

with its two differential invariants

$$\phi_1=\frac{u_1}{u^{3/2}},\qquad \phi_2=\frac{u_2}{u^2},$$

where

$$u=(r+1)y^{(r)}y^{(r+2)}-(r+2)(y^{(r+1)})^2,$$

$$u_1=(r+1)^2(y^{(r)})^2y^{(r+1)}-3(r+1)(r+3)y^{(r)}y^{(r+1)}y^{(r+2)}+2(r+2)(r+3)(y^{(r+1)})^3,$$

$$u_2=3(r+4)(r+3)(r+2)(y^{(r+1)})^4-6(r+4)(r+3)(r+1)y^{(r)}[y^{(r+1)}]^2y^{(r+2)}+4(r+4)(r+1)^2(y^{(r)})^2y^{(r+1)}y^{(r+3)}-(r+1)^3(y^{(r)})^3y^{(r+4)}$$

does not admit the equation $F=0$. For this group, as well as for the groups of Articles 22 to 26, the equations derived to determine A, B and C are inconsistent.

THE THEORY OF TESTIMONY

First Paper: A Single Witness

By J. S. TAYLOR

The traditional theory of testimony has been severely criticised by J. M. Keynes in his "Treatise on Probability."[1] He says, in part, "It may, however, be safely said that the principal conclusions on the subject set out by Condorcet, Laplace, Poisson, Cournot, and Boole, are demonstrably false. The interest of the discussion is chiefly due to the memory of these distinguished failures." While this criticism may be somewhat merited, a rather detailed investigation of the subject has led the author of the present paper to the conclusion that the criticism is really directed against the lack of explicit statement of restrictions which the authors cited may have considered as implied and the lack of clear definition of terms whose meaning they may have thought self-evident. It would hardly seem, therefore, that their principal conclusions are "demonstrably false," but rather that the lack of explicit statement referred to above has resulted in considerable confusion. It is in the hope that this confusion may be in some small part relieved that the following discussion is presented.

The problem of the theory of testimony is the problem of determining the probability of the occurrence of an event when such occurrence is attested to by one or more witnesses of known "credibility." Confusion in the discussion of such problems has arisen chiefly on two grounds. First, because of ambiguity in the phrase "credibility of the witness"; and, second, because of lack of explicit statement concerning the probable distribution of incorrect observations when the witness is in error. The first is a matter of definition, but lack of explicit definition would seem to be universal. The second involves the question of the nature of error and necessitates either an *a priori* assumption concerning the nature of error or an *a posteriori* conclusion based

[1] J. M. Keynes, "A Treatise on Probability," p. 180 (Macmillan, 1921).

on experiment concerning the probable distribution of the incorrect observations of a given observer in a given experiment.

In addition, as Keynes has pointed out, confusion has also arisen in some instances concerning the question as to whether the problem was that of determining the ratio of the number of times the observer (or observers) correctly stated the event occurred to the number of times such statement was made; or the ratio of the number of times the observer (or observers) correctly stated the event occurred to the number of times it actually did occur. The first is the probability that the event occurred if the observer (or observers) said it did; the second is the probability that the observer (or observers) said the event occurred if it did occur. We assume the former to be the problem.

In this first paper the discussion will be limited to problems in which there is a single witness, leaving the theory of the combination of testimonies for discussion in a second paper. It will be assumed throughout that the witness does not try to deceive or, at least, that his tendency to err, whatever its cause, obeys the general law embodied in Postulate 1, which is assumed to hold in all the problems discussed in this paper.

This postulate needs, perhaps, a little introductory explanation. If the theory of testimony is to mean anything, it must be assumed that there is some quantitative measure of an observer's reliability which is invariant under a change of experiment to whose results he attests; that while the weight which we attach to his testimony may well vary with the type of experiment whose results he records there is nevertheless some quantitative measure of accuracy depending upon the witness alone which is the same for all experiments, or at least for all experiments of a large class. In order to arrive at such a quantitative measure we shall assume that the cause of error is something in the nature of periodic temporary blindness; that when an experiment is repeated a large number of times a witness accurately observes and records the result that occurs α times in N on the average and that β times in N (with $\alpha+\beta=N$) his faculties fail to *observe* the event that occurred and that he records at randon any one of the results he knows to be possible, the correct result included. And it is further assumed that the ratio of α to N and of β to N is the

same for a given observer for all experiments (or for all experiments of a large class with the experiments to be discussed all belonging to one class). More briefly stated these assumptions are embodied in the following postulate:

Postulate 1. *Random Error. If an experiment is repeated a large number of times a given observer accurately records the result that occurs* α *times in* N, *on the average, and the remaining* β *times in* N *records any one of the known possible results at random, with the ratio of* α *to* N *and of* β *to* N *the same for all experiments.*

Thus a witness may record the correct result of an experiment either by reason of his accuracy or by accident, the probability of the latter depending largely upon the experiment itself. The probability that the result he records is correct, due to either of these causes, will be called the " credibility of the witness " in recording that result of that experiment.

Definition 1. *If in recording the occurrence of a certain one of* k *possible results of an experiment which is repeated a large number of times the ratio of the number of times a given observer records such occurrence correctly to the number of times he so records is* P_i, P_i *will be called the " credibility " of the given observer in recording this result of this experiment;* $(1-P_i)$ *will be called his " incredibility."*

The " credibility " of a witness varies, then, in different experiments and in recording different results of the same experiment, since the number of times he records a given result correctly includes those times when he records that result by chance. In fact, the " credibility " of a witness as defined is the probability that an event occurred if he so attests.

Definition 2. *The " personal accuracy " of observation,* P, *of a given observer is the limit of his credibility in attesting to the occurrence of one of* k *different possible results of an experiment as* k *approaches infinity; his " personal inaccuracy " is* $1-P$.

Theorem 1. *If Postulate* 1 *holds,* P *is unique for a given observer.* For as the number of possible results of an experiment approaches infinity the probability that the observer will record

by chance the result that occurs approaches zero, and the probability that he records the correct result approaches the ratio of the α and N of Postulate 1, which was there assumed to be the same for all experiments.

Now consider an observer of known personal accuracy of observation P. Suppose one ball drawn from a bowl containing k balls numbered from 1 to k, the number of the ball drawn recorded by an observer, the ball replaced, and this process repeated a large number of times. If there were an infinite number of differently numbered balls the observer would be right α times in N and wrong β times in N, where $\frac{\alpha}{N} = P$ and $\frac{\beta}{N} = 1-P$. By Postulate 1 this means that β times in N the observer records the number of any one of the balls at random. Consider kN drawings. Of this number ball No. i will be drawn N times (on the average). α of these N times the observer will record No. i by reason of his personal accuracy and $\frac{\beta}{k}$ times by chance (it is assumed the two numbers β and N with ratio equal to $1-P$ have been taken so that β is divisible by k). In the other $(k-1)N$ drawings, in which ball No. i was not drawn, the observer records No. i by chance $\frac{(k-1)\beta}{k}$ times. So that the ratio of the number of times the observer records No. i correctly to the number of times he records No. i is $P_i = \frac{\alpha + \frac{1}{k}\beta}{\alpha+\beta}$. The same reasoning gives the following theorem.

Theorem 2. *If Postulate* 1 *holds and if the personal accuracy of an observer is* P, *then his credibility in recording the occurrence of any one of* k *equally probable events is*

$$P_i = \frac{1}{k} + \frac{k-1}{k}P = P + \frac{1-P}{k};$$

and his incredibility is

$$1-P_i = \left(1-\frac{1}{k}\right) - \frac{k-1}{k}P = (1-P) - \frac{1-P}{k}.$$

As before stated P_i is the probability that the given event occurred if the witness says it did. It may be noted that as $\frac{1}{k}$, the *à priori* probability of the occurrence of the event, approaches 1, or certainty, P_i approaches 1; and that as $\frac{1}{k}$ approaches zero P_i approaches P. Also, the increase in the probability of the occurrence of the event due to the testimony of the witness is the product of his " personal accuracy " and the *à priori* probability of the non-occurrence of the event; the increase in the reliability of his testimony due to the *à priori* probability of the occurrence of the event (the increase of P_i over P) is the product of his " personal inaccuracy " and the *à priori* probability of the occurrence of the event. These results are in seeming disaccord with those heretofore obtained as will be pointed out in discussion of the more general case where the results of an experiment are not all equally probable.

The equations of Theorem 2 are readily solved for P and $1-P$ in terms of P_i and $1-P_i$, giving.

Theorem 3. *If Postulate* 1 *holds and if in recording the result of an experiment having* k *equally probable outcomes* $(k>1)$, *the credibility of a given observer is* P_i *and his incredibility is* $1-P_i$, *then his personal accuracy is*

$$P=P_i-\frac{1-P_i}{k-1},$$

and his personal inaccuracy is

$$1-P=(1-P_i)+\frac{1-P_i}{k-1}.$$

The necessity of assuming $k>1$ is obvious if we note that if k equals 1 the credibility of the witness is certainty and hence no clue whatever concerning his personal accuracy is obtainable.

Now if the credibility of an observer in an experiment all of whose possible outcomes are equally probable is known, his personal accuracy can be found, and, hence, his credibility in any other experiment of the same type can be determined. We thus have

Theorem 4. *If Postulate* 1 *holds and if the credibility of an observer in recording the results of an experiment having k equally probable outcomes is* P_i (*where* $k>1$), *then the observer's credibility in recording the results of an experiment having k' equally probable outcomes is*

$$P_j = P_i + (1-P_i)\frac{\frac{1}{k'}-\frac{1}{k}}{1-\frac{1}{k}}.$$

Next let us consider cases where the k possible results are not all equally likely. Take the classical problem of a bowl containing 1,000 balls 999 of which are black and the remaining one white. Suppose an observer whose personal accuracy is $P=\frac{\alpha}{N}$, and whose personal inaccuracy is $1-P=\frac{\beta}{N}$, records drawings as before. Suppose, also, that he knows the bowl contains only black balls and white balls but is ignorant of their relative numbers; suppose, in fact, all of the assumptions of Postulate 1 hold. Then, by reason of his personal accuracy he will record the correct color α times in N and will record one of the two colors at random β times in N. Suppose $1{,}000N$ drawings are made. Then of the $999N$ times that black is drawn he will record black 999α times by reason of his personal accuracy and $999\frac{\beta}{2}$ times by chance. Also, of the N times white is drawn he will record black $\frac{\beta}{2}$ times by chance. Hence he will record black $999\alpha+500\beta$ times and of those he will record black correctly $999\alpha+999\frac{\beta}{2}$ times. Hence, since the probability that black was drawn if the witness records black is the ratio of the number of times he records black correctly to the number of times he records black, this probability is

$$P_b = \frac{999\alpha+999\frac{\beta}{2}}{999\alpha+500\beta} = \frac{999+999P}{1{,}000+998P},$$

which is, by definition of " credibility," the observer's credibility in recording this result of this experiment; and the probability that he is in error when he records black is

$$1-P_b=\frac{\frac{\beta}{2}}{999\alpha+500\beta}=\frac{1-P}{1{,}000+998P}.$$

Also, his credibility in recording white (the probability that white was drawn if he so attests) and his incredibility in recording white (the probability that white was not drawn if he records white) are, respectively,

$$P_w=\frac{\alpha+\frac{\beta}{2}}{\alpha+500\beta}=\frac{1+P}{1{,}000-998P}$$

and

$$1-P_w=\frac{\frac{999\beta}{2}}{\alpha+500\beta}=\frac{999-999P}{1{,}000-998P}.$$

It may be noted here (as other results have indicated, although there was once wide divergence of opinion) how largely the *à priori* probability of the occurrence of an event affects an observer's credibility. The above results are not, however, in accord with any of the solutions of this problem the author has been able to find in the literature. Laplace,[2] for example, concludes that if the " witness deceives one time in ten " the probability that the white ball was drawn if he so attests is $\frac{9}{1008}$. But investigation of preceding discussions shows that by the statement the " witness deceives one time in ten " Laplace meant neither that the " credibility " of the witnessis $\frac{9}{10}$ nor that his " personal accuracy " is $\frac{9}{10}$, in the sense these terms have been here employed. In his discussion of the drawing of a ball from an urn containing a thousand differently numbered balls, he concludes that the probability that a given ball was

[2] Laplace, " Philosophical Essay on Probabilities," pp. 109–117 (English translation of sixth edition, by Truscott & Emory, Wiley & Sons, 1902).

drawn if the witness so attests is the veracity of the witness itself, which would seem to indicate that he defines "veracity of the witness" to mean what has here been called his "credibility"; but he exhibits this result as a conclusion, not as a definition.[3] Furthermore, while no such explanation is explicitly given, Laplace's reasoning in the discussion of the black and white ball problem just discussed leads to the conclusion that by the "veracity of the witness" he means the ratio of the number of times the witness correctly states the event occurred to the number of times the event actually did occur (the probability that the witness said the event occurred if it did occur), and that this ratio is the same for all experiments. Granted this meaning and this assumption Laplace's result is correct; in short, starting with the solution of one of the two problems which, as stated in the introduction, have sometimes been confused, Laplace correctly obtained the answer to the other one. Whether or not the assumption that the veracity of the witness as defined by implication by Laplace is the same for all experiments is any less in accord with fact than Postulate 1 of the present paper is, perhaps, well open to question. Some such assumption would seem to be desirable if any generalization is to be hoped for; the author can only point out that these two different assumptions are not consistent one with the other.

Lacroix,[4] in obtaining the same result as that given by Laplace for the black and white ball problem, would seem to be using the same definition of "veracity of the witness" as that implied by Laplace. But Lacroix exhibits the solution of this problem as an illustration of the principal credited to Condorcet of considering the *à priori* probability of the occurrence of an event as the testimony of a second witness and using the result obtained for the probability of the occurrence of an event when two independent witnesses agree that it occurred. Now it is this result

[3] As indicated at the end of the discussion of Theorem 2, the results there obtained are not in accord with the traditional theory. Laplace concludes that if the veracity of a witness is less than the *à priori* probability of the occurrence of an event, his testimony decreases the probability of the occurrence of the event. The difference in these results is obviously due to difference in definition of the phrase "veracity of the witness."

[4] Lacroix, "Calcul des Probabilités" p. 225 and pp. 230–232 (4me édition, Mallet-Bachelier, Paris, 1864).

concerning the combined testimony of two witnesses which has been pointed to in particular by Keynes as a "demonstrably false" conclusion obtained with practical universality by all writers on the theory of testimony. While a discussion of the justice of this criticism is beyond the confines of this first paper it may at least be stated that the formula referred to is not correct, in general, if the definition of veracity credited to Laplace is employed in its derivation.

In conclusion of this discussion it should be mentioned that our results apparently do not agree with Keynes. Keynes calls attention to a method of regarding a "witnesses's credibility not absolutely, but with reference to a given type of question," proposed by W. E. Johnson. Keynes' statement reads, "If a represents the fact of A's testimony regarding x, then we may define A's credibility for x as α, where α is given by the equation

$$x/ah = x/h + \alpha\sqrt{x/h.\bar{x}/h};$$

so that $\alpha\sqrt{x/h.\bar{x}/h}$ measures the amount by which A's assertion of x increases its probability." Since x/h and $\bar{x}/h$ are the *à priori* probabilities of the occurrence and non-occurrence of the event, respectively, the latter part of this quoted statement would not seem consistent with the explanation on page 181 that a/a_1h "is the probability that a statement is true of which we know only that X_1 has asserted it," for x/ah would thus seem to mean both the probability that a statement is true of which we know only that A has asserted it and at the same time the *à priori* probability of the occurrence of x plus "the amount by which A's assertion of x increases its probability." Moreover, as practically no other explanation than that outlined is given no clue is given as to why this "definition" seemed desirable or as to what it actually signifies.

In continuation of the discussion of the black and white ball problem consider next the more general case of a bowl containing k balls, k_1 of which are marked No. 1, k_2 marked No. 2, . . . , and k_r marked No. r, with $k_1+k_2+\ \ldots\ +k_r=k$. Suppose kN drawings of a single ball as before (ball replaced after each drawing). Then, in the long run, ball No. i would be drawn k_iN times in kN drawings. If the observer's personal accuracy is

$P=\frac{\alpha}{N}$ and his personal inaccuracy is $1-P=\frac{\beta}{N}$, he will in kN drawings record No. i correctly $k_i\alpha$ times by reason of his personal accuracy and $k_i\frac{\beta}{r}$ times by chance. Also, he will record No. i incorrectly $k_1\frac{\beta}{r}+k_2\frac{\beta}{r}+\ \ldots\ +k_r\frac{\beta}{r}-k_i\frac{\beta}{r}=(k-k_i)\frac{\beta}{r}$ times. Hence,

Theorem 5. *If Postulate* 1 *holds and if an observer whose probable personal accuracy is* P *records the results of an experiment having* r *possible results of which the à priori probabilities are* $p_1, p_2, \ldots, p_r$, *respectively, where* $p_1+p_2+\ \ldots\ +p_r=1$, *then, if he records the occurrence of the* ith *result, the probability that the* ith *result actually occurred is*

$$P_i=p_i+p_i\frac{rP(1-p_i)}{(rp_i-1)P+1}$$ [5]

If $p_i=\frac{1}{k}$ and $r=k$ this result reduces to that of Theorem 2. It is to be noted that (rp_i-1) may be either positive or negative depending on the problem, but that $(rp_i-1)P+1$ will always be positive, giving $P_i>Pi$, as it should. Also, the larger p_i is the smaller the numerator of the fraction and the larger the denominator; so that, the larger p_i is the smaller is the *percentage* increase of probable occurrence due to the testimony of the witness.

As before, we may solve for P in terms of P_i and p_i, giving

Theorem 6. *If Postulate* 1 *holds and if in recording the results of an experiment having* r *possible outcomes of which the à priori probabilities are* $p_1, p_2, \ldots, p_r$, *the credibility of a given wit-*

[5] Here, as in the following theorems, it is assumed p_1, p_2, etc., are rational numbers; that the theorem applies to experiments which can be put into one to one correspondence with an experiment of drawing balls from a bowl, which includes practically all types of importance in which it could be hoped Postulate 1 holds.

ness in recording the occurrence of the ith result is P_i, then his personal accuracy of observation is

$$P=\frac{P_i-p_i}{P_i-p_i+rp_i(1-P_i)}$$

Now since the assumption of random error independent of the number of possible outcomes and the *à priori* probabilities of those outcomes means that P is unique for a given observer in all experiments where Postulate 1 holds, the above result should be the same not only for all values of i but for all experiments where Postulate 1 holds.

Thus $\frac{P_i-p_i}{(P_i-p_i)+rp_i(1-P_i)}=\frac{P_j-p_j}{(P_j-p_j)+r'p_j(1-P_j)}$. This gives

Theorem 7. If Postulate 1 holds and if in recording the results of an experiment having r possible outcomes the probability that a given result of à priori probability p_i occurred when such occurrence is attested to by a given observer is P_i, and the probability that any other result of the same or different experiment having r' possible outcomes ($r'=r$ if same experiment) of à priori probability p_j occurred when the same observer so attests is P_j, then

$$\frac{P_i-p_i}{P_j-p_j}=\left(\frac{r}{r'}\right)\left(\frac{p_i}{p_j}\right)\left(\frac{1-P_i}{1-P_j}\right);$$

or, the ratio of the increases in probability of occurrence due to the witness's testimony is the product of the ratio of the numbers of possible results, the ratio of the à priori probabilities, and the ratio of the probabilties that the testimonies are in error.

This result offers a partial check on the validity of Postulate 1. If the accuracy of testimony in recording the occurrence of two or more events can be determined (the observer's credibility for each of the events) then the above result must be satisfied if random error as assumed exists.

Either the successive application of Theorem 5 or solution of the equation of Theorem 7 gives

Theorem 8. If Postulate 1 holds, and if in recording the results of an experiment having r possible outcomes the probability that a

given result of à priori probability p_i occurred when such occurrence is recorded by a given observer is P_i, then the probability that any other result of à priori probability p_j of the same or different experiment having r' possible outcomes ($r'=r$ if the experiments are the same) occurred when such occurrence is recorded by the same observer is

$$P_j = p_j \frac{P_i(r'-rp_i)+(r-r')p_i}{P_i(r'p_j-rp_i)+(r-r'p_j)p_i};$$

if $r'=r$

$$P_j = p_j \frac{P_i(1-p_i)}{P_i(p_j-p_i)+p_i(1-p_j)}.$$

NOTES ON DYNAMICAL SYSTEMS NON-INTEGRABLE BY SEPARATION OF VARIABLES AND ON THE EXISTENCE OF "UNMECHANICAL" ORBITS IN THE ATOM

By M. SANDOVAL VALLARTA

1. There has been some speculation lately as to the existence of holonomic, conservative, reversible dynamical systems whose equations of motion cannot be integrated by a separation of variables in the Hamilton-Jacobi equation and a good deal has been written about "unmechanical" orbits in the atom. The purpose of this note is twofold: First, two examples in classical mechanics are given showing that, either by constraining a mass-point to move freely in a given type of coördinate space, or else by introducing arbitrary anisotropic fields of force, dynamical systems which are not integrable by separation of variables can be constructed. The question as to the *physical* significance of such systems is left for the present entirely untouched. Second, an attempt is made to show that, while conclusions as to the existence of unmechanical orbits in atomic systems, drawn from a failure to obtain correct values of the energy terms through the calculus of perturbations do not seem to be wholly justified, arguments of a more general character based on the theory of relativity point out that in fact such is the case: unmechanical orbits do exist in the atom.

2. **Examples of Dynamical Systems non-Integrable by Separation of Variables.** We consider this question from the standpoint of classical dynamics. Restricting our considerations for the sake of simplicity to systems with *two degrees of freedom*, in which case coördinate-space reduces to a surface, Stäckel and Levi-Civita have shown[1] that the *only* types of the first fundamental form $ds^2 = a_{ij}dx_i dx_j (i, j = 1, 2)$ which lead to an expression

[1] Mathematische Annalen, Vol. 59, 392, 1904. Cf. also the author's "Note on the Quantization of Non-Conditioned-Periodic Systems" in the preceding number of the Jour. of Math. and Phys.

for the kinetic energy T in which the variables can be separated are:

(1) Plane fundamental form referred to parametric curves which are lines of translation.

(2) Fundamental form corresponding to a surface applicable on a surface of revolution, for which $x_1 = \text{const.}$ are the parallels.

(3) First fundamental form of Liouville's type

$$ds^2 = [\,U(u) + V(v)\,](du^2 + dv^2)$$

which for $U(u)$ or $V(v)$ constant may reduce to (1) or (2).

It follows, in the first place, that the Hamilton-Jacobi equation of a point moving under no forces, *i.e.*, the equation determining the geodesics on the corresponding surface, is integrable by separation of variables (by quadratures) when, and only when, the mass-point moves freely upon a Liouville surface, upon a plane, or upon a surface of revolution. It is well known that if a mass-point moves upon any one of these surfaces under the action of a field of force, a separation of variables may still be possible, for example, when it moves subject to two gravitational centers.[2] But in any case the variables chosen to express the potential energy must be such as to give a fundamental form of one of the three types above, otherwise the kinetic energy is made non-separable. Thus, in the above example, the elliptic coördinates which make the system separable give a linear element of Liouville type

$$ds^2 = (\cosh^2\xi - \cos^2\eta)(d\xi^2 + d\eta^2).$$

It is therefore seen that the form of the linear element conditions the separability of that part of the Hamiltonian function which corresponds to the kinetic energy and in turn restricts the variables to be used in the expression for the potential energy. Hence the conclusions: (1) The potential energy can be expressed only in those systems of variables which make the kinetic energy separable and, (2) the kinetic energy is separable if, and only if, the mass point is moving upon a surface which admits a first fundamental form of the types specified above.

[2] Jacobi, Vorlesungen über Dynamik, p. 190, 198. Reimer, Berlin, 1884.

3. Furthermore, it follows, perfectly in general, that any mass-point which moves upon a surface which *is not of any one of the three types specified above under the action of any forces whatsoever*, constitutes a dynamical system which is not integrable by separation of variables. Since an infinite number of such surfaces are known to exist, an infinite number of non-separable (non-conditioned-periodic) systems also exist.[3]

As an example of a dynamical system which is made non-separable by a suitable anisotropic field of force, take the case of a point moving on a Liouville surface, with a potential-energy function

$$\overline{V} = -1/2\,[k_1 f_1(u) g_1(v) + k_2 f_2(u) g_2(v)\,]$$

k_1, k_2 being constants ($k_1 \neq k_2$) and the f's and g's arbitrary functions having finite continuous first and second derivatives. The kinetic energy is

$$T = 1/2\,[\,U(u) + V(v)\,](\dot{u}^2 + \dot{v}^2).$$

Hence the Hamilton function is

$$H(p, q) = \frac{1}{2\,[\,U(u) + V(v)\,]}\,(p_u^2 + p_v^2) + \overline{V}$$

and the Hamilton-Jacobi equation is

$$\left(\frac{dW}{du}\right)^2 + \left(\frac{dW}{dv}\right)^2 - [\,U(u) + V(v)\,][k_1 f_1(u) g_1(v) + k_2 f_2(u) g_2(v)\,] = 2\,[\,U(u) + V(v)\,]E$$

E being the energy constant. Owing to the presence of the two constants, the variables cannot be separated whatever be our arbitrary functions, as is seen readily, even though somewhat laboriously, by substituting in the Levi-Civita equations of condition.[4]

[3] For a discussion of the conditions under which a surface is of Liouville type, see for example Bianchi, Lezioni di Geometria differenziale, p. 303, Pisa, 1922. Criteria for determining without integration whether a given surface is applicable on a Liouville surface have been developed by Ricci, Lezioni sulla teoria della superficie, Chap. VII., Padova, 1898. The above conclusion can be reached in a variety of ways, see for example Darboux, Leçons sur la théorie générale des surfaces, Vol. II, p. 208, Paris, 1889.

[4] See reference footnote (1) or the author's paper quoted therein.

4. **Unmechanical orbits in the atom.** In a recent remarkable paper, Born and Heisenberg[5] have undertaken a systematic investigation of all possible types of quantized paths in the excited helium atom by means of the perturbation method of Born and Pauli[6] and come to the very interesting conclusion that "a consecutive quantum-theoretical treatment of the helium problem leads to incorrect values for the energy-terms" and further that "obviously there are but two ways out of the difficulty; either the quantum conditions are false — or else the motion of the electrons in the stationary states is not governed by the equations of classical mechanics." From this standpoint, it is equally obvious, as they point out, that whichever alternative is chosen, new fundamental difficulties arise for the understanding of those quantum-theoretical results which have been confirmed by experiment, notably the Stark effect and the normal Zeeman effect. To these remarks may be added the following: (1) No method of quantization of a perturbed motion which is wholly free from objection has yet been devised and no single example of a successful application of the calculus of perturbations to quantum-theoretical atomic problems leading to *positive* results has yet been given. To be sure, the calculus of perturbations has been used to *disprove* previous atomic models, as in the well-known paper of Kramers[7] on Kemble's helium model, to quote but a single instance. In a more general way, the nature of the difficulties involved is beautifully illustrated in a recent note of P. Ehrenfest.[8] This consistent failure is perhaps what has led most writers to conclude the existence of "unmechanical" orbits in atomic systems: In this sense the "failure of classical mechanics" is a proven fact. But in view of the objections which can be made to any of the proposed methods of applying the calculus of perturbations to the quantization of atomic systems,

[5] Zeitschrift für Physik, Vol. 16, p. 229, 1923.
[6] Zeitschrift für Physik, Vol. 10, p. 137, 1922.
[7] Zeitschrift für Physik, Vol. 13, p. 312, 1923.
[8] Zeitschrift für Physik, Vol. 19, p. 242, 1923, cf. also N. Bohr Annalen der Physik, Vol. 71, p. 277, footnote, 1923. It may be noted in connection with Ehrenfest's note and the author's paper quoted above that doubts as to the sufficiency of Fermi's proof that a normal mechanical system is quasi-ergodic have been expressed by W. Urbanski, Physikalische Zeitschrift, Vol. 25, p. 47, 1924.

such conclusions are not justified except on plausibility grounds. It must be borne in mind that the atom with more than one electron is a perturbed system not at all analogous to the perturbed astronomical system, for the reason that the presence of commensurable periodic times in the unperturbed motion is exceptional in the astronomical case and the probability for its occurrence may be taken to be infinitely small, whereas such commensurabilities must occur in the motion of electrons in the atom, on account of the very nature of the quantum conditions, which demand exact rational ratios of periodic times in the unperturbed orbits.[9]

(2) In the absence of any proof to the contrary, it can hardly be expected that the quantization of a dynamical system in variables chosen at random will give the correct values of the energy-terms. Whether the helium atom is a dynamical system which is integrable by separation of variables or not, we do not now know; but even if it were shown that incorrect values of the energy-terms are obtained by its quantization in a system of separable variables, the conclusion need not be that the helium atom is a " non-classic " mechanical system but rather that our quantum rules, which, as is well known, are of a wholly heuristic nature, still need revision. Difficulties of a similar nature occur in the theory of the complex Zeeman effect[10] but in view of the semi-empirical nature of this theory, no conclusions of a far-reaching character seem to be justified. It should be strongly emphasized that the above remarks are not intended to implicate that an atom with more than one electron *is* a classical dynamic system — arguments to the contrary will be soon brought up — but rather to point out that, from the standpoint of the calculus of perturbations, the question cannot yet be as definitely answered as seems to be usually supposed.

5. From the standpoint of the Theory of Relativity, the problem of the existence of unmechanical orbits in the atom can

[9] Born & Heisenberg, Zeitschrift für Physik, Vol. 14, p. 45, 1923; cf. also L. Nordheim, ibid, Vol. 17, p. 316, 1923.
[10] A. Landé, Physikalische Zeitschrift, Vol. 24, p. 441, 1923.

be more definitely answered. In a material field[11] Hamilton's equations are symmetrical only in the particular case of a static field, as clearly brought out in a recent paper by Mecke.[12] On the other hand, the Lagrangian equations conserve their form, as would appear obvious *a priori* owing to their derivation directly from the (generalized) Hamilton principle.

This may be directly shown as follows (cf. Mecke, l.c.): The total action is the sum of field and substance-action of electricity plus the substance-action of mass.[13] The substance-action of electricity is

$$(S_e) = -\int de \int \phi_k dx_k$$

where ϕ_k is the electrodynamic potential, the outer integration is over all substance-elements and the inner along that part of the world-line described within the world-region considered by the substance-element with the charge de. The field being defined by the electromagnetic tensor F_{ik},

$$F_{ik} = \frac{\partial \phi_k}{x_i} - \frac{\partial \phi_i}{\partial x_k},$$

its action is in the same notation, by the Maxwell-Lorentz equations,

$$(S_f) = \tfrac{1}{4}\int F_{ik} F^{ik} dV.$$

Lastly, the substance-action of mass is

$$(S_m) = \int dm \int \sqrt{-g_{\mu\nu} dx_\mu dx_\nu},$$

where again the outer integration is over all substance-elements, the inner along that part of the world-line of an arbitrary substance-element of mass dm which lies within the world-region considered.

The total action in a material field is therefore

$$S = \int dm \int \sqrt{-g_{\mu\nu} dx_\mu dx_\nu} - \int de \int \phi_k dx_k + \tfrac{1}{4}\int F_{ik} F^{ik} dv.$$

[11] A material field is one obtained by the superposition of an electromagnetic on a massic field. Cf. T. De Donder. La gravifique einsteinienne, Ch. VII, p. 80, Paris, 1921.

[12] R. Mecke, Zeitschrift für Physik, Vol. 21, p. 26, 1924.

[13] H. Weyl. Raum, Zeit, Materie, p. 195, Berlin, 1923.

In the case of an electron, this reduces to:

$$S=\int(m_0c\sqrt{-g_{\mu\nu}dx_\mu dx_\nu}+e\phi_k u^k)ds=\int L(x_k,\ u^k)ds$$

where $u^k=\dfrac{(dx)^k}{ds}$ is the "velocity" (Eddington) — or more properly perhaps the "world direction k" (Weyl). The Hamilton principle

$$\delta S=0$$

yields at once the Lagrangian equations:

$$\frac{d}{ds}\left(\frac{\partial L}{\partial u^i}\right)-\frac{\partial L}{\partial x_i}=0 \quad (i=1,\ 2,\ 3,\ 4)$$

$P_i=\dfrac{\partial L}{\partial u^i}$ is the momentum. Further, remembering that

$$\sqrt{-g_{\mu\nu}u^\mu u^\nu}=1,$$

it is readily shown that (De Donder, l.c.)

$$H(P_i,\ u^i)=P_i u^i$$
$$m_0cu^i=-g^{ik}(P_k-e\phi_k).$$

From which are derived,

$$\frac{dx_k}{ds}=\frac{\partial H(P_k,\ u^k)}{\partial P_k}=\frac{\partial H(P_k,\ x_k)}{\partial P_k}+\frac{P^k}{m_0c}$$

and again it can be readily shown, as done by Mecke (l.c.), that these can be reduced to the canonic equations:

$$\frac{dq_k}{dt}=\frac{\partial W}{\partial p_k},\qquad \frac{dp_k}{dt}=-\frac{\partial W}{\partial q_k},$$

only if the field is static, *i.e.*, independent of the time. Hence we obtain the important conclusion: From the standpoint of the theory of relativity, unmechanical orbits, *i.e.*, orbits not obeying the Hamilton canonic equations, and hence not obtainable through Hamilton-Jacobi's method, may exist in the atom, for the latter's

field is *in general* a non-static material field.[14] A theoretical reason for the existence of such orbits, often emphasized by Bohr and others, is, therefore, found. There is a possibility, however, that satisfactory approximations may be found through the method conceived by Smekal and discussed by the author in a previous paper (l.c., footnote 1).

The above conclusion, however, is to a certain extent weakened by the fact that the quantum conditions, in order to possess any usefulness, must again be purely heuristic. Mecke enunciates the quantum postulate in a perfectly general form as follows: The line-integral of the substance-action $\int L ds$ over a closed world-line is always an integral multiple of the element of action, Planck's h. Now, the solution of our four Lagrangian equations contains, in general, eight integration constants, of which four are determined by the "initial" conditions, while four remain arbitrary. In order that the system may be quantized, four quantum conditions are thus necessary. For this purpose, Mecke *assumes* that the substance-action can be split up by separation of variables into four conditions

$$I_k = \int P_k dx_k = n_k h \qquad k = 1, 2, 3, 4$$

where P_k is the generalized impulse, $P_k = \dfrac{\partial L}{\partial \dot{w}^k}$. The similarity between this and Sommerfeld's well-known assumption is obvious. Again the quantum conditions are seen to be purely heuristic, so long as no proof of the existence of a system of separable variables is available. Besides, Mecke's formulation of the quantum conditions imposes a restriction on the choice of coördinates which is wholly foreign to the principle of relativity. To this and other related questions, reference will be made in a forthcoming paper.

[14] The material field of the hydrogen atom is, at least approximately, static and has spherical symmetry. This point will be fully discussed in a forthcoming paper. Cf. Mecke, l.c., also Jaffé, Annalen der Physik, Vol. 67, p. 212, 1922.

Contribution Number 3,
The Research Laboratory of Theoretical Physics,
March, 1924.

NOTE ON A PROPERTY OF RECTILINEAR LINES OF PRINCIPAL STRESS

By Paul Heymans

Maxwell[1] showed that in the two dimensional stress problem, for elastic isotropic bodies, obeying the law of linear proportionality between stress and strain, and allowing the cubical dilatation to be substituted for the sum of the three principal linear dilatations, the lines of equal inclination of principal stresses and the loci of the points of equal algebraic difference of magnitudes of the principal stresses completely determine the stress distribution and should suffice to yield at all points the directions and the magnitudes of the two principal stresses.

The temporary double refraction due to stress, observed in transparent bodies optically isotropic in the unstressed state, offers a means for obtaining these isoclinic lines and the points of equal algebraic difference of the principal stresses.

Attempts to make these optical data yield the separate values of the principal stresses had until lately remained unsuccessful.

In the photoelastic method[2], the isochromatic lines were supplemented by the determination of the sum of the principal stresses, which in the two dimensional elastic stress problem can be readily shown to be proportional to the deformation in a direction normal to the plane of the two principal stresses.

From the equations of equilibrium of a curved elementary rectangle bounded by four near-by lines of principal stress, derived by Mesnager[3], in the case of two-dimensional stress and on the assumption that no body forces exist, L. N. G. Filon[4] developed

[1] J. Clerk Maxwell, On the Equilibrium of Elastic Solids. Trans. Royal Soc. Edinburgh, 1850, Vol. XX, Part I.

[2] Paul Heymans, La Photo-Elasticimétrie, ses principes, ses méthodes et ses applications. Bull. Soc. Belge Ing. et Ind., II, 2, 1921.
Also, Trans. Am. Soc. Mech. Eng., 1922, and Tech Engineering News, June, 1922.

[3] M. Mesnager, Equilibre élastique des plaques circulaires. Annales des Ponts et Chaussées, 1909, IV.

[4] L. N. G. Filon, On the Graphical Determination of Stress from Photo-elastic Observations. Engineering, Oct. 19, 1923.

a method of graphical integration by successive steps, allowing the derivation of the separate values of the principal stresses throughout the stressed body from the isoclinic and isochromatic lines.

These equations of equilibrium of the curved elementary rectangle are:

$$\left.\begin{aligned} \frac{\partial p}{\partial s_1} + \frac{p-q}{\rho_2} &= 0 \\ \frac{\partial q}{\partial s_2} + \frac{p-q}{\rho_1} &= 0 \end{aligned}\right\} \quad \text{(I)}$$

where p and q are the two principal stresses, s_1 and s_2 arcs taken along the lines of principal stress corresponding to the stress p and q respectively, ρ_1 and ρ_2 the radii of curvature of the two lines of principal stress[8].

The definite integrals of the equations (I) give the following two equations:

$$\left.\begin{aligned} p &= p_o - \int_o^A \frac{p-q}{\rho_2} ds_1 \\ q &= q_o - \int_o^A \frac{p-q}{\rho_1} ds_2 \end{aligned}\right\} \quad \text{(II)}$$

where p_o and q_o are the values of p and q at the lower limit of the interval of integration.

In equations (II), whenever the radius of curvature of one of the lines of principal stress becomes infinite, the corresponding integral vanishes. *Hence the two following propositions:*

1. *Throughout any elementary surface band which is limited by two nearby rectilinear parallel lines of principal stress;*

2. *In any section traversed normally or obliquely by rectilinear parallel lines of principal stress;*

The principal stress normal to the latter is constant in direction and in magnitude.

[8] The lines of principal stress, in general determined by the intersection of the isostatic surfaces, can be readily derived analytically or graphically from the isoclinic lines, for the two-dimensional stress.

If it vanishes at any point, it vanishes throughout the surface or the section.

These propositions lead, in many problems of stress distribution, to valuable *a priori* results.

The following may serve as an illustrative example:

Consider a section A_1A_n of a plate where the lines of principal stress are as given in Fig. 1.

Across the section A_1A_n, the lines of principal stress are parallel to the longitudinal axis.

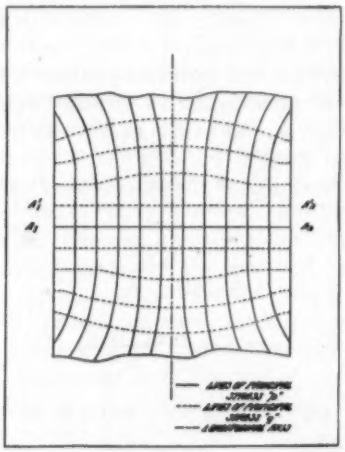


Fig. 1.

At either point A_1 or A_n, which lies on the free boundary, the transverse principal stress "q" is zero. At all points along A_1A_n the radius of curvature of the line of principal stress p is equal to ∞.

Hence: the transverse stress "q", orthogonal to "p", is:

$$q = q_o - \int_o^A \frac{p-q}{\infty} ds_2 = q_o - 0 \equiv 0$$

i.e., zero *throughout* the section.

Also: across all sections $A_1'A_n'$, defined by a line of principal stress "q" parallel to A_1A_n, the stress distribution is the same as across the primitive section A_1A_n.

Contribution Number 4,
The Research Laboratory of Theoretical Physics,
Massachusetts Institute of Technology,
Cambridge, Mass.,
March 19, 1924.

NOTE ON THE CURL

By W. H. INGRAM

The usual approach to a definition of the *curl* in textbooks on vector analysis is via the Hamiltonian operator ∇. A better way[1], in that a physical interpretation is correlated with the definition, is as follows: Consider the vector equation of motion of a point P of a non-rigid body carrying along with it a coördinate system with origin at another point of the body O. The motion of the point P at a particular instant is given by

$$v = v_t + \omega \times \mathbf{r} + v' \tag{1}$$

where v_t is the translational motion of the origin O, ω the angular velocity of the coördinate system, $\mathbf{r}$ the position-vector OP (the first two terms of the right-hand side of (1) thus giving the rigid-body motion), and v' the motion arising purely out of the non-rigidity of the body.

When r is an infinitesimal, the velocity of the point P is also given by

$$v = v_t + \mathbf{r} \cdot \text{grad} v \tag{2}$$

so that

$$v' = \mathbf{r} \cdot \text{grad}\ v - \omega \times \mathbf{r} \tag{3}$$

Now, the equation (3) expressed in analytics exhibits v' as a linear vector function of $\mathbf{r}$ so that from a simple property of linear operators[2], the analysis of v' into a *symmetric* linear vector function and a second term (a vector product) is a natural step. The symmetric matrix of differential coefficients obtained in this way represents the *dilatation* of v. The second term, from the symmetry of its components is recognizable by the student as a vector which is obtained in some way from v'. He is thus led

[1] The latest book on mathematical physics, whose author has also written a Vector Analysis, contains the fundamental error of using the curl of the velocity vector to give an analytic expression of angular velocity for substitution into fluid-body equations.

[2] See Sec. 168 of Heaviside's Electromagnetic Theory.

to the obvious name and definition of the *curl* and finds incidentally that

$$v = v_t + \phi\mathbf{r} + \tfrac{1}{2}[(\text{curl } v) \times \mathbf{r}] \tag{4}$$

where $\phi\mathbf{r}$ is a linear symmetric vector function of $\mathbf{r}$.

From equations (1) and (4) it is seen that the difference in the motion of a non-rigid body to that of a rigid body is made up of elastic-body extensions and shears given by $\phi\mathbf{r}$ and the motion $[\frac{1}{2}(\text{curl } v) - \omega] \times \mathbf{r}$ due to the positional variation of the angular velocity.

NOTE ON A METHOD OF EVALUATING THE COMPLEX ROOTS OF A QUARTIC EQUATION

By W. V. Lyon

The numerical solution of any physical problem which involves a linear differential equation with constant coefficients depends upon our ability to solve a corresponding algebraic equation. It often happens that the solution of the differential equation is in the form of a damped sinusoid, in which case the roots of the algebraic equation are complex. If these complex roots are represented by vectors, a relatively simple method may be developed for the solution of a fourth-degree equation. Consider the general fourth-degree equation

$$a+b\mathbf{x}+c\mathbf{x}^2+d\mathbf{x}^3+x^4=0. \qquad (1)$$

The variable $\mathbf{x}$ is printed in Clarendon type to indicate that it is a vector. A common vector notation among engineers is $\mathbf{x}=\underline{x/\theta}$. Thus (1) may be written

$$a+b\underline{x/\theta}+c\underline{x^2/2\theta}+d\underline{x^3/3\theta}+\underline{x^4/4\theta}=0. \qquad (2)$$

The horizontal and vertical components of these five vectors with respect to a line through $\underline{cx^2/2\theta}$ must be zero, and thus we may write:

$$cx^2+(bx+dx^3)\cos\theta+(a+x^4)\cos 2\theta=0 \qquad (3)$$

and

$$(bx-dx^3)\sin\theta+(a-x^4)\sin 2\theta=0. \qquad (4)$$

Since $\sin\theta$ is not zero, it may be factored from equation (4) giving:

$$\cos\theta=-\tfrac{1}{2}\frac{bx-dx^3}{a-x^4}. \qquad (5)$$

If this value of $\cos\theta$ is substituted in equation (3) it becomes, on writing y in place x^2:

$$a^3-a^2cy+(abd-a^2)y^2+(2ac-ad^2-b^2)y^3+(bd-a)y^4-cy^5+y^6=0. \qquad (6)$$

Since y represents the square of the *magnitude* of $\mathbf{x}$ there are but two values in which we are interested. They may be considered as the roots of

$$a-my+y^2=0 \tag{7}$$

in which m is yet to be determined from the coefficients of equation (6).

Let the other factor of (6) be:

$$a^2+gy+hy^2+ky^3+y^4=0. \tag{8}$$

Multiply (7) and (8) and equate the coefficients of like powers of y in the result and in equation (6). This process gives four equations involving the four unknown coefficients m, g, h, and k.

$$\begin{aligned} a^2m+ag &= -a^2c \\ gm+a^2+ah &= abd-a^2 \\ g+mh+ak &= 2ac-a^2d-b^2 \\ k+m &= -c. \end{aligned}$$

If all but m are eliminated there remains the condition that:

$$m^3-cm^2+(bd-4a)m-ad^2-b^2+4ac=0. \tag{9}$$

The solution for y from equation (7) is:

$$y=\frac{m\pm\sqrt{m^2-4a}}{2}.$$

Since y is a real positive number m must be greater than $2\sqrt{a}$. The method of solution is now clear. Find the one real positive root of Equation (9) which is greater than $2\sqrt{a}$ by any method that seems desirable,— the " pinch method," for example. Substitute this in Equation (7) and solve for y. The numerical values of x are the square roots of these two values of y. For each value of x there is an associated angle which is determined by Equation (5). Letting θ_1 and θ_2 be these angles, we have as the value of the roots of Equation (1):

$$x_1\ (\cos\ \theta_1 \pm j \sin\ \theta_1)$$

and

$$x_2\ (\cos\ \theta_2 \pm j \sin\ \theta_2).$$

There follows a numerical example:

$$290-18x+47x^2-6x^3+x^4=0.$$

The auxiliary cubic is

$$m^3-47m^2-1052m+43756=0.$$

The root of this cubic which is greater than $2\sqrt{290}$ is 50.75 (using a 20-inch slide rule).

$$y=\frac{50.75\pm\sqrt{50.75^2-4\times290}}{2}$$

$$=44.20 \text{ or } 6.65$$

$$x_1=6.65 \quad x_2=2.56$$

$$\cos\ \theta_1=-\tfrac{1}{2}\frac{-18+6\times6.65^2}{290-6.65^4}\,6.65=0.494$$

$$\sin\ \theta_1=0.870$$

$$\cos\ \theta_2=-\tfrac{1}{2}\frac{-18+6\times2.56^2}{290-2.56^2}\,2.56=-0.111$$

$$\sin\ \theta_2=.994$$

$$\therefore\ x=3.29\pm j\ 5.78$$

$$\text{or}=-0.283\pm j\ 2.55$$

ON THE ATOMIC MODELS OF THE ALKALI METALS[1]

By Robert Bruce Lindsay

TABLE OF CONTENTS

		Page
	Introduction	191
I.	Method of Calculation	197
II.	The Lithium Atom	207
III.	The Sodium Atom	208
IV.	The Potassium Atom	212
V.	The Rubidium Atom	215
VI.	The Caesium Atom	218
VII.	Summary of Sections I–VI with Discussion of Results	221
VIII.	Diagrams of the Orbits	226
IX.	Fixed Phase Relations in the Sodium Atom	230
X.	The Polarization Effect in Lithium and Sodium	232
	Acknowledgments	236

Introduction. In a recent work Bohr[2] has outlined a theory of atomic structure according to which in the normal state of the atom the electrons are divided into groups corresponding to a classification by means of quantum numbers. One of the principal features of this classification is that the orbit of every electron is characterized in a way analogous to that in which the stationary states of a particle moving in a central field of force are fixed by two conditions, which may be written in the form

$$J=nh, \qquad P=kh \tag{1}$$

where h is Planck's constant, while n and k are two integers, the so-called principal quantum number and azimuthal quantum

[1] This memoir was submitted in partial fulfillment of the requirement for the degree of Doctor of Philosophy from the Massachusetts Institute of Technology in April, 1924.

[2] N. Bohr, Three essays on the theory of spectra and atomic constitution, Camb. Univ. Press, 1922; cf. also N. Bohr, Ann. d. Phys. 71, 1923, hereinafter referred to as loc. cit.

number respectively. P represents the angular momentum of the electron round the center of the orbit multiplied by 2π, while J may be defined by the relation

$$J - P = \int p_r dr \tag{2}$$

where r is the length of the radius vector, and p_r the radial momentum, and where the integral has to be taken over a complete period of the radial motion.

In the limiting case where the field of force in which the electron moves differs very little from the attraction of a central nucleus with a charge τe, the orbit will differ very little from a slowly rotating Kepler ellipse, the major axis $2a$ and parameter $2l$ (*latus rectum*) of which are given by the relations:

$$2a = \frac{n^2h^2}{2\pi^2\tau e^2 m} \tag{3}$$

$$2l = \frac{k^2h^2}{2\pi^2\tau e^2 m} . \tag{4}$$

In this case the energy W necessary to remove the electron to infinite distance will, to a close approximation, depend only on the principal quantum number and will be given by

$$W = \frac{2\pi^2\tau^2 e^4 m}{h^2 n^2} . \tag{5}$$

This type of orbit occurs, as is well known, in the stationary states involved in the emission of series spectra where one electron moves during its whole revolution in the region outside the other electrons, and where accordingly in the formulæ above we have to introduce for τ the total number of unit charges in the atomic residue. In the normal state of the atom we also meet with such orbits when we consider inner regions where the attraction of the nucleus outweighs the repulsion of the other electrons. In this case the orbital dimensions and energy will be given approximately by the formulæ above if for τ we introduce the atomic number N. In general, however, the central orbit of an electron in the atom will differ much from such a slowly rotating Keplerian orbit, and the energy will depend essentially both on n and k. In a large number of cases the central orbit described by the electron will

still exhibit certain simple properties. In fact the orbit can often be considered as a sequence of loops, each coinciding closely with a part of a Keplerian orbit around a point charge which may be referred to as the effective nuclear charge N^*e. This charge is equal to the nuclear charge minus the charge on the electrons which move in orbits of smaller dimensions, and which to a first approximation may be said to screen each a unit positive charge on the nucleus. These outer loops are linked to inner loops where the electron penetrates into the region of the orbits of the inner electrons where the screening effect of these electrons is considerably diminished. In these parts the orbits will deviate essentially from Keplerian orbits, and in particular the closest distance to which the electron approaches the nucleus will, in general, be much smaller than the perihelion distance of the ellipse of which the outer loop forms a part. While the parameter (*latus rectum*) of this latter ellipse will be given by the relation (4) by putting $\tau = N^*$, the major axis of the ellipse, as well as the energy necessary to remove the electron from the atom, will *not* be given by the formulæ (3) and (5), if we set $\tau = N^*$. We may, however, introduce an effective quantum number n^* for the orbit, which is defined in such a way that the major axis and work under consideration are given approximately by

$$2a = \frac{n^{*2}h^2}{2\pi^2 N^* e^2 m} \tag{6}$$

$$W = \frac{2\pi^2 N^* e^4 m}{h^2 n^{*2}} . \tag{7}$$

Now Bohr has shown that the remarkable periodicity in the properties of the elements, brought to light by the periodic table, may be accounted for by the circumstance that although in successive periods of the table we have to do with increasing quantum numbers for the outermost electrons in the atom, the orbits of these electrons will nevertheless be characterized by effective quantum numbers which differ only a little from each other. A striking example of this circumstance is afforded by the alkali metals, which form the first natural family in the periodic system and for the neutral atoms of which the following constitution is assumed:

TABLE I

Z \ n k	1_1	2_1	2_2	3_1	3_2	3_3	4_1	4_2	4_3	4_4	5_1	5_2	5_3	5_4	5_5	6_1
Li 3	2	1														
Na 11	2	4	4	1												
K 19	2	4	4	4	4		1									
Rb 37	2	4	4	6	6	6	4	4			1					
Cs 55	2	4	4	6	6	6	6	6	6		4	4				1

N denotes atomic number; the various types of orbits are indicated in the top row, while the numbers in the table indicate the number of orbits of each type.

To provide a better understanding of the preceding table, it will be well for us to consider the recent Bohr theory of atomic structure for a moment in greater detail than has been done in the introduction just given. The Bohr conception of atomic constitution is that of the successive capture and binding of the individual electrons in the field of force about the nucleus. During the capture each individual electron passes through a series of stationary states emitting radiation and finally drops into a normal orbit — its final resting place in the normal, unexcited state of the atom — characterized by the quantum numbers n and k previously mentioned (page 191). A review of the working of this process may best be gained by an examination of the complete table of orbital types (Table 2) of which Table 1 is an excerpt.

Without going into too great detail, we shall follow the building process for the first few elements of the periodic system. From the foregoing, it is clear that the final, normal state for a single electron moving in the field of force due to a nucleus of charge e will be a circular orbit whose diameter is given by Equation (3) if we make the substitution $n=1$, and $\tau=1$. This orbit will, by definition, be a 1_1 orbit. The result, of course, is the simple hydrogen atom first discussed by Bohr[3] in 1913. Coming next

[3] N. Bohr: On the Constitution of Atoms and Molecules, Phil. Mag. 26 1–25, 476–502, 857–875 (1913).

TABLE 2. ELECTRON ORBIT TYPES OF THE ELEMENTS

N \ n_k	1_1	2_1	2_2	3_1	3_2	3_3	4_1	4_2	4_3	4_4	5_1	5_2	5_3	5_4	5_5	6_1	6_2	6_3	6_4	6_5	6_6	7_1	7_2
1 H	1																						
2 He	2																						
3 Li	2	1																					
4 Be	2	2																					
5 B	2	2	(1)																				
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
10 Ne	2	4	4																				
11 Na	2	4	4	1																			
12 Mg	2	4	4	2																			
13 Al	2	4	4	2	1																		
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
18 A	2	4	4	4	4																		
19 K	2	4	4	4	4		1																
20 Ca	2	4	4	4	4		2																
21 Sc	2	4	4	4	4	1	(2)																
22 Ti	2	4	4	4	4	2	(2)																
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
29 Cu	2	4	4	6	6	6	1																
30 Zn	2	4	4	6	6	6	2																
31 Ga	2	4	4	6	6	6	2	1															
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
36 Kr	2	4	4	6	6	6	4	4															
37 Rb	2	4	4	6	6	6	4	4			1												
38 Sr	2	4	4	6	6	6	4	4			2												
39 Y	2	4	4	6	6	6	4	4	1		(2)												
40 Zr	2	4	4	6	6	6	4	4	2		(2)												
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
47 Ag	2	4	4	6	6	6	6	6	6		1												
48 Cd	2	4	4	6	6	6	6	6	6		2												
49 In	2	4	4	6	6	6	6	6	6		2	1											
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
54 X	2	4	4	6	6	6	6	6	6		4	4											
55 Cs	2	4	4	6	6	6	6	6	6		4	4				1							
56 Ba	2	4	4	6	6	6	6	6	6		4	4				2							
57 La	2	4	4	6	6	6	6	6	6		4	4	1			(2)							
58 Ce	2	4	4	6	6	6	6	6	6	1	4	4	1			(2)							
59 Pr	2	4	4	6	6	6	6	6	6	2	4	4	1			(2)							
.	.	.	.	.	.	.	.	.	.	.	.	.	.			.							
71 Lu	2	4	4	6	6	6	8	8	8	8	4	4	1			(2)							
72 .	2	4	4	6	6	6	8	8	8	8	4	4	2			(2)							
.	.	.	.	.	.	.	.	.	.	.	.	.	.			.							
79 Au	2	4	4	6	6	6	8	8	8	8	6	6	6			1							
80 Hg	2	4	4	6	6	6	8	8	8	8	6	6	6			2							
81 Tl	2	4	4	6	6	6	8	8	8	8	6	6	6			2	1						
.	.	.	.	.	.	.	.	.	.	.	.	.	.			.	.						
86 Nt	2	4	4	6	6	6	8	8	8	8	6	6	6			4	4						
87 .	2	4	4	6	6	6	8	8	8	8	6	6	6			4	4					1	
88 Ra	2	4	4	6	6	6	8	8	8	8	6	6	6			4	4					2	
89 Ac	2	4	4	6	6	6	8	8	8	8	6	6	6			4	4	1				(2)	
90 Th	2	4	4	6	6	6	8	8	8	8	6	6	6			4	4	2				(2)	
.	.	.	.	.	.	.	.	.	.	.	.	.	.			.	.	.				.	
118 ?	2	4	4	6	6	6	8	8	8	8	8	8	8	8	8	6	6	6				4	4

to the binding of the second electron when the nuclear charge is $2e$, the study of the helium spectrum (strictly speaking, the so-called parhelium spectrum) indicates that this electron will also be bound in a 1_1 orbit, whose plane, however, makes an angle of 60° with the plane of the electron first bound. The complete study of the helium atom is complicated by the fact that in addition to the ordinary stable helium (parhelium) there exists a metastable modification (as shown by the researches of Franck[4] and his colleagues), the so-called orthohelium, the spark spectrum of which indicates that its second electron is captured in an orbit which corresponds to an energy only $\frac{1}{6}$ of the energy of binding of the second electron in the normal state. Such an orbit will, on the Bohr theory, be a 2_1 orbit. The normal (parhelium) state will, however, surely contain two electrons both bound in 1_1 orbits.

Bohr now assumes that in the formation of all atoms of atomic number greater than two the first two electrons are bound in 1_1 orbits similar to those in helium, though more firmly held of course because of the greater nuclear field. The capture of the third electron will explain the structure of lithium, the first of the alkali metals, and the method of its binding is read from the lithium arc spectrum, where it is found that the final normal orbit of this electron must be a 2_1 orbit, since binding in a 1_1 orbit would be wholly unable to account for the series spectrum.

It is unnecessary to follow further the details of the building process. In each case the addition of an electron is studied from the observation of the spectrum of the atom thus formed. Where parentheses enclose a number in the table, doubt is indicated as to the exactness of the choice of orbit. The cases of Boron, Scandium and others, for example, demand further spectroscopic evidence before the uncertainty may be removed completely. The main outlines of the table must, however, in the light of Bohr's fundamental assumptions, be regarded as correct. Examination shows that in the main the development of the atom takes place in the building up of the outer groups of electrons as, for

[4] Franck and Reiche: ZS. f. Phys. 21, 635 (1920).

example, in the third period of the periodic system in the formation of the atoms of the elements from Neon to Argon. This development of the outer groups is of great significance from the standpoint of the difference in the series spectra of the successive elements and those physical properties which depend essentially on what happens at the periphery of the atom[6]. As we go further down the periodic table, however, another type of development becomes evident. For beginning with Scandium ($N=21$), we meet a building up of the inner orbital groups while that of the outer groups temporarily ceases. Bohr thus makes clear why in this part of the table we first encounter *successive* elements with remarkably similar ordinary physical properties. As an example, we may call attention to the famous Iron triad of ferromagnetic metals. Other such groups are similarly explained. In a general way, then, it has been possible on the basis of the Bohr conception of atomic formation to give an interpretation of atomic characteristics in general, as well as the phenomena of spectral line emission and absorption.

It is highly desirable, however, to be able ultimately to attach quantitative as well as qualitative significance to the orbital types present in the various atoms, that is, to determine, with suitable additional and as far as possible *simple* assumptions, the actual dimensions of the orbits symbolized by the quantum numbers in the table. By virtue of the relatively large amount of data now available for the alkali elements, this family provides a very interesting point of attack. It will then be the object of this dissertation to calculate on the basis of certain simple assumptions the dimensions of the various orbits present in the atoms of the alkali elements, and where possible to discuss the results with reference to those obtainable from other sources.

I. **Method of Calculation.** The calculations leading to the results presented in this paper are based on the simplifying assumption that each electron in the atom will influence the orbits of the others as a continuous distribution of electricity, the density of which is uniform over every spherical surface around the nucleus and is of such magnitude that the total charge between

[6] See A. Sommerfeld, Atombau und Spectrallinien, 3d Ed., page 120 ff.

two such surfaces is equal to the fraction of the charge of the electron corresponding to the fraction of the time which the electron spends between these surfaces during its revolution. The following diagram (Fig. 1) will serve to illustrate the assumption

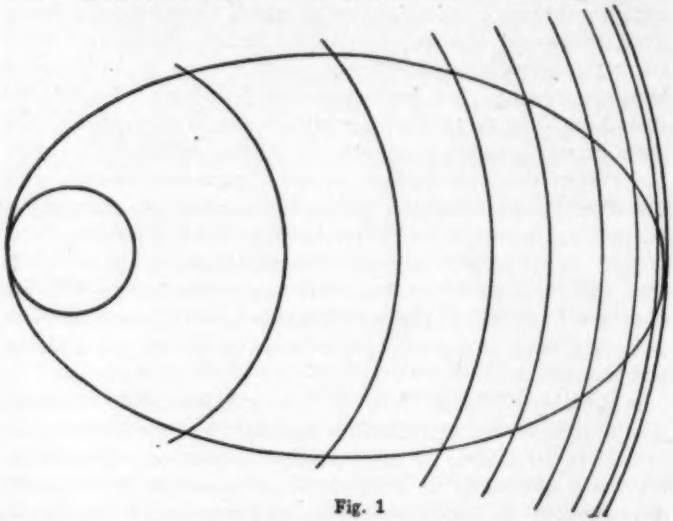


Fig. 1

just stated. The elliptic orbit in the figure is considered, as far as its influence on other orbits is concerned, as replaced by a series of spherical shells, with radii depending on the time distribution of the electron in the orbit. In the diagram in question each shell save the innermost and outermost is supposed to have a charge equivalent to $e/8$ where e is the charge on the electron. The innermost and outermost shells have a charge of $e/16$. The radius of the outermost shell is, of course, equal to the aphelial distance of the electron in the elliptic orbit. The radius of the succeeding one is equal to the distance of the electron from the nucleus when it has traveled toward the nucleus for a time $T/16$ where T is the period of revolution of the electron in its orbit. For the case of an elliptic orbit, we easily determine this distance from the equation

$$\omega t = u - \epsilon \sin u \qquad (8)$$

where $\omega = 2\pi/T$, t is the time, ϵ is the eccentricity of the ellipse, and u the eccentric anomaly. When ω and ϵ are known, u may be found graphically (the equation being transcendental) for any value of t. It must be emphasized here that all this holds for elliptic orbits only, and serves only as an illustration. For the distribution in the case of the actual, non-Keplerian orbits encountered in our work, we refer to the discussion below.

In the special case where the orbits are circular and described by the same quantum numbers, the above assumption will be interpreted to mean that the mutual screening effect of each such electron is taken simply to be $e/2$ charges. This is, of course, the mean value of the effect of a charge e immediately inside and immediately outside the shell on which it moves.

The above assumption is the simplest rational one at present available not only for the interaction of electrons in groups characterized by different quantum numbers, but also for the interplay of electrons moving in orbits characterized by the same quantum number. In this latter case any attempt to introduce more detailed assumptions regarding the relative orientations of the orbits might be at present of very doubtful value. Indeed we should have to consider rotation of the loops of the orbits and the precession of the orbital planes as well as the phase relations between the electrons in the various orbits of the same quantum number. For these last relations, as has been pointed out by Bohr, we cannot assume any simple kind of polygonal or polyhedral symmetry, but shall have to consider phase relations of a more complex character. In their effect the latter may not differ greatly from the above assumption. On the other hand it must be emphasized that our assumption in its summary form does not provide for certain effects regarding the mutual perturbations of the orbits which play an important part in the interpretation of optical spectra. These effects, having to do with the complex structure of the spectral lines, will, however, not be dealt with in this work. A general discussion of them may be found in the paper of Bohr's first mentioned. We may note here, however, that the question of phase relations for electrons in elliptic orbits has been treated to some extent in the writings of several authors.

For example Sommerfeld[6] has examined the case in which the assumption is made that all the electrons in a given sub-group move so as to visit the vicinity of the nucleus simultaneously. This provides what might be called a pulsating electron system. Close examination shows, however, that this scheme breaks down in any attempted explanation of X-ray spectra. Another possible way of attacking the problem and probably the only other simple way is the one suggested at one time by Bohr, according to which the electrons in a given sub-group of non-circular orbits attain their perihelia successively at equal intervals of time. Thus in the case of four electrons (the four 2_1 orbits in sodium, for example), the electrons will visit the nucleus at successive intervals equal to $T/4$ where T is the period of revolution of each electron. We may represent the situation schematically in the following diagram (Fig. 2), assuming for the sake of simplicity of presentation

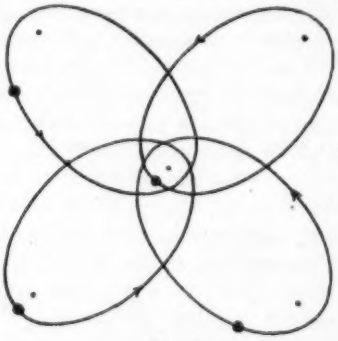


Fig. 2

that the orbits are all in the same plane, with their major axes oriented perpendicularly to each other, and neglecting the rotation effect. While the principal results of the present research have been obtained without consideration of phase relations, the structure of the sodium atom has been worked out using the hypothesis suggested immediately above. The theory and results will be found in Section X.

The matter of phase relations leads one further to discuss at a little greater length the question of the orientation of the orbital

[6] A. Sommerfeld, loc. cit. page 612.

planes. Bohr[7] has assumed that the planes of the four 2_1 orbits in the carbon atom (a very stable atom) are so related that they are normal to the lines joining the center to the vertices of a regular tetrahedron. This so-called tetrahedral symmetry of the 2_1 orbits would, however, probably have to be modified considerably in the cases of neon and sodium and the higher alkali atoms, due to the disturbing influence of the 2_2 orbits. The rotation of the orbits in their planes and the precession of the orbital planes demanded by the experimental evidence of the complex structure of spectral lines also help to complicate in a most disconcerting way any attempt to introduce a simple geometric orientation of these planes. While work on this point has not been attempted in this research, such a consideration will probably prove highly desirable and yield most interesting results. (See last paragraph of this section.)

In the calculation of the orbital dimensions under our assumption the following general method based on the quantum theory relations is used. Let V denote the potential energy of the electron and T its kinetic energy. Then if p_r and p_ϕ denote the radial and angular momenta of the electron respectively, we have the relation

$$T = \frac{1}{2m}\left(p_r^2 + \frac{p_\phi^2}{r^2}\right) \tag{9}$$

where m is the mass of the electron and r its distance from the nucleus. If E denotes the total energy of the electron we have accordingly

$$E = V + \frac{1}{2m}\left(p_r^2 + \frac{p_\phi^2}{r^2}\right) \tag{10}$$

or

$$p_r^2 = 2mE - 2mV - \frac{p_\phi^2}{r^2}\,. \tag{11}$$

If we assume a central field, V will be a function of r only and p_ϕ will remain constant during the motion. In the stationary states p_ϕ, according to (1) will be given by

$$p_\phi = \frac{kh}{2\pi} \tag{12}$$

[7] N. Bohr, Three essays on the theory of spectra and atomic constitution, p. 93.

where k is the angular quantum number. Substitution of these results into the radial quantum condition yields according to (1) and (2)

$$\int p_r dr = \int \sqrt{2mE - 2mV - \frac{k^2h^2}{4\pi^2r^2}}\, dr = (n-k)h \qquad (13)$$

where n is the principal quantum number and the integration is conducted over a complete circuit from the smaller of the two roots of the expression under the radical to the larger and back again. For convenience we shall call this expression Q.

For orbits of a given type the substitution of the appropriate value of E and the potential energy function V would yield upon integration the proper integral value for $n-k$. In the actual problem confronted, however, the appropriate E and the course of V are unknown to begin with. In fact, on our assumption, V will depend on the electron orbits. We are then forced to proceed by approximations. For the latter it is of great significance to note that in calculating the field we can for a first approximation use Keplerian ellipses instead of the actual orbits. The reason for this is to be found in the fact that for the larger part of each period the electrons move in the outer loops of their orbits where the motion is nearly Keplerian. For the same reason it will be seen that the groups of orbits of larger dimensions will have but little influence on the orbits of smaller dimensions. The necessary procedure in fixing the orbits will therefore be to begin with the groups of lower quantum number and for each type of orbit to take for a first approximation what seem to be reasonable values for the effective nuclear charge and for the effective quantum number corresponding to the outer orbital loop. This will then give a value of the energy E and the dimensions of the corresponding Keplerian ellipses from which V can be calculated according to our assumption. The Q function may then be plotted as a function of r and the indicated integration carried out. The deviation of the result from the appropriate integral value of $n-k$ for the type of orbit in question will then indicate the direction in which the preliminary assumption regarding the orbit was wrong. The procedure now is by successive approximations to find the value of E which satisfies (13) and then after calculating

the corresponding orbit (which will, of course, be a non-Keplerian orbit) to readjust the function V and again to seek a value for E satisfying (13).

It will be worth while at this point to consider rather carefully the fixation of the orbits satisfying these conditions. This will involve a brief discussion of the Hamiltonian theory of varying action[8]. By definition the *action* of a particle moving under the influence of a given field of force is the line integral of the momentum of the particle taken between the limits of its motion. Thus if mv denotes the momentum of the particle and r_0 and r_1 denote the limits of the motion, we have

$$A = \int_{r_0}^{r_1} mv\,dr. \tag{14}$$

The ordinary use of this quantity is connected with the celebrated principle of "least action," according to which if a particle moves freely or under forces such as occur in nature between any two points, the value of the action for the actual path is less than that along any other conceivable path. More correctly, for the actual path, we have $\delta A = 0$, *i.e.*, the action is *stationary*.

For our present problem, however, we must use action in the sense which Hamilton denoted as "varying." Let us consider again Equations (9) and (10) which may be written

$$T = \tfrac{1}{2}mv^2 = E - V. \tag{15}$$

If in the variation of v the *initial* velocity of the particle varies, then E must vary, and the general expression for the variation of the action in such a case, where also we assume that the initial and final coördinates vary, is therefore

$$\begin{aligned}\delta A = {} & m[\dot{x}\delta x + \dot{y}\delta y + \dot{z}\delta z] \\ & - m[(\dot{x})_0\delta x_0 + (\dot{y})_0\delta y_0 + (\dot{z})_0\delta z_0] + t\delta E\end{aligned} \tag{16}$$

where $\dot{x} = \dfrac{dx}{dt}$, etc., t is the time and the subscript zero indicates the initial state. Without loss of generality as far as our work

[8] See Tait and Steele, Dynamics of a Particle, Seventh Edition, page 275 ff. Also A. Sommerfeld; loc. cit. page 674 ff.

is concerned, we may consider $\delta x_0 = \delta y_0 = \delta z_0 = 0$ and have simply

$$\delta A = m[\dot{x}\delta x + \dot{y}\delta y + \dot{z}\delta z] + t\delta E. \tag{17}$$

Now if A could be found as a function of x, y, z (*i.e.*, the final coördinates) and E, the first and second integrals of the motion would obviously take the form

$$\frac{\partial A}{\partial x} = m\dot{x}, \quad \frac{\partial A}{\partial y} = m\dot{y}, \quad \frac{\partial A}{\partial z} = m\dot{z} \tag{18}$$

together with

$$\frac{\partial A}{\partial E} = t. \tag{19}$$

The determination of the motion in a given case then reduces to the solution of the partial differential equation

$$\left(\frac{\partial A}{\partial x}\right)^2 + \left(\frac{\partial A}{\partial y}\right)^2 + \left(\frac{\partial A}{\partial z}\right)^2 = 2m(E - V). \tag{20}$$

Let us suppose that we have a solution of this equation in the form $A = f(x, y, z, a, b, E)$ where a and b are two integration constants or parameters. Then it can be shown that the geometrical equation of the orbit of the particle will be given in the form

$$\frac{\partial A}{\partial a} = C_1, \qquad \frac{\partial A}{\partial b} = C_2 \tag{21}$$

while the time equation of the orbit will be given by

$$\frac{\partial A}{\partial E} = t + C_3 \tag{22}$$

where C_1, C_2, and C_3 are three additional constants.

We now apply these results to the special case at hand. Since we assume the field to be central, the Differential Equation (20) will become in polar coördinates r, ϕ

$$\left(\frac{\partial A}{\partial r}\right)^2 + \frac{1}{r^2}\left(\frac{\partial A}{\partial \phi}\right)^2 = 2m(E - V). \tag{23}$$

The solution obtained by " separation of the variables " is in the form

$$\left.\begin{aligned} \frac{\partial A}{\partial \phi} &= \text{constant} = a \\ \left(\frac{\partial A}{\partial r}\right)^2 &= 2m(E-V) - \frac{a^2}{r^2} \end{aligned}\right\} \tag{24}$$

Hence we have at once

$$A = a\phi + \int \sqrt{2m(E-V) - \frac{a^2}{r^2}}\, dr. \tag{25}$$

From Equations (21) and (22) we have for the geometrical equation of the orbit

$$\frac{\partial A}{\partial a} = \phi - a \int \frac{dr}{r^2 \sqrt{2m(E-V) - \frac{a^2}{r^2}}} = C_1 \tag{26}$$

and for the time equation of the orbit

$$\frac{\partial A}{\partial E} = \int \frac{m dr}{\sqrt{2m(E-V) - \frac{a^2}{r^2}}} = t + C_2. \tag{27}$$

It will be noted that the expression under the radical in (26) and (27) is identical with Q previously defined in Equation (13), since $a = \frac{\partial A}{\partial \phi} = p_\phi = \frac{kh}{2\pi}$ under the assumption of a central field. To find, then, the actual orbit or rather the second approximation to the orbit, in any given case we have only to take that Q which when used in Equation (13) gave good quantization, and use it to evaluate the integrals in (26) and (27). Thus from (27) we deduce at once the new spherical distribution corresponding to the second approximation to the orbit. For it is seen that Equation (27) is the generalized expression of which Equation (8) is a special case. We may also determine at the same time the fraction of its period which the electron moving in this new and more nearly correct orbit spends inside any shell of circular orbits. As has been indicated above, the procedure then is to use this

second approximation to recalculate V and Q and again to compute the quantum integral (13), working by successive approximations until we have an orbit which is not only quantized (*i.e.*, satisfies (13)) but which is also consistent in that it yields on analysis by our assumptions the same field values and the same Q by which it was quantized. It might be suspected at first that this process of successive approximations would be extremely tedious. It is found, however, that the influence of the shape of the orbit on the course of V and hence of Q is relatively small when once we have an approximately quantized orbit, and so in general, two or three approximations are completely sufficient to yield consistent, quantized orbits.

It should be noted here that in the evaluation of the integral in (27) we make use of the fact previously emphasized that in both the extreme inner and extreme outer parts of the orbit, the motion is nearly Keplerian. Therefore from r_{min} to $r_{min}+\eta$, and from $r_{max}-\eta'$ to r_{max} (where η and η' are small relative to r_{max}) we calculate the integral in (27) using the expression (8). This is done to avoid the graphical integration up to an infinite discontinuity, since the integration in (27) must be carried out between the two roots of Q.

Although it is the main object of this research to calculate only the *dimensions* (*i.e.*, aphelial distances) of the orbits of the various types present in the atoms of the alkali metals, it is clear that the Equation (26) yields us the equation in polar coördinates of each orbit considered. In Section VIII there will be found drawings of the orbits obtained in this way, with explanatory material.

At this point reference should be made to the interesting work of Fues[9], who has shown in the case of the sodium atom how it is possible to derive the energies necessary to remove the various electrons in the atom as they are given by the experimental results in optical and X-ray spectra by assuming a suitable central field in which the electrons move. His work is thus largely empirical in nature, differing therein from that of the present research, for the latter attempts to account for the intra-atomic fields of force by a direct consideration of the orbits of the elec-

[9] E. Fues, ZS. f. Phys. 11, 364, 1922; 12, 1, 1922; 13, 211, 1923.

trons to which this field is due. Other interesting work in the same field is that done by D. R. Hartree[10], whose investigations to some extent parallel those of Fues, in that he has calculated the orbits in the atoms of sodium and potassium directly from the energy values obtained from analysis of their spectra. As in the case of Fues's work, that of Hartree is essentially the inverse process of the one used in the present research. In his discussion, however, he takes up the question of orbital plane orientation and makes an interesting application of his results to data on X-ray scattering experiments. In Section VII of this dissertation, comparison is made of his results with those of the present research. We may perhaps take this opportunity to mention also the investigation of Van Urk[11] who has extended the preliminary work of Schrödinger[12] on the calculation of the effective quantum numbers of non-circular outer electron orbits which penetrate inner shells. This latter work, however, pays no attention to the dynamical character of the interplay of the electronic orbits of the inner as well as the outer groups, and therefore the results obtained, while interesting, are of somewhat formal nature.

II. **The Lithium Atom** ($N=3$). The discussion of the lithium atom by the present method is very simple. For the group of 1_1 orbits the mutual screening effect will on our assumption be 0.5 units[13] and the effective nuclear charge is consequently $3-0.5=2.5$ units. The radius of these orbits is therefore 0.40 a_0 where a_0 is the radius of the first quantum orbit in hydrogen, and equals 0.532×10^{-8} cm. We next note that the electron in the 2_1 orbit will, on our assumption, perform its motion wholly outside the 1_1 orbit. Indeed the perihelion distance for a 2_1 orbit in a field of unit charge will be 0.54 a_0 or greater than the radius of the 1_1 orbits. This means that on our assumption the outer orbit will be a simple Keplerian ellipse with effective quantum number 2. The actual effective quantum number for this orbit as given by

[10] D. R. Hartree, Proc. Camb. Phil. Soc., Vol. XXI, part 6 (1923); see also: Phil. Mag. December, 1923.

[11] A. Th. van Urk, ZS. f. Phys., 13, 268, 1923.

[12] I. Schrödinger, ZS. f. Phys. 4, 347, 1921.

[13] The unit charge is, of course, the charge e, and will hereafter be so understood.

the spectral term[14] is, however, 1.59. This result shows at once the limitations of our method of procedure. Our assumption in fact here leads to the result that the arc spectrum of lithium should not differ from the hydrogen spectrum. In this connection reference may be made to the above-mentioned paper by Bohr in which it is pointed out that the deviation of the effective quantum number of the 2_1 orbit in lithium from 2 is due largely if not entirely to the polarization induced in the inner configuration by the outer electron. This effect will be neglected in the calculations leading to the principal results of this research. It is of such importance, however, that it merits some further attention, and therefore a preliminary investigation of it for the case of the atoms of lithium and sodium has been carried out. This will be found with accompanying calculations in Section X. It will there be seen that in a very simple way we can account for the above noted discrepancy between our value of n^* and that given by the spectral term.

Our interest at this point in the lithium atom is simply to facilitate comparison with the results obtained in the case of the other alkali metals on the same general assumption, *i.e.*, that enunciated in Section I.

III. **The Sodium Atom** ($N=11$). In the case of sodium the fact that the inner orbits are almost wholly independent of the outer ones makes it possible to fix at once the 1_1 orbits with a considerable degree of approximation. Assuming a mutual screening effect of 0.5 unit, we obtain for the radius of these orbits $a_0/10.5=0.095a_0$.

In considering next the groups of 2-quantum orbits the first problem concerns the relative positions of the 2_1 and 2_2 orbits.

[14] It will be recalled (see N. Bohr, loc. cit., page 241) that the Rydberg spectral series formula may be written in the form

$$\nu = \frac{K}{(n^*)^2} - \frac{K}{(n'^*)^2}$$

where ν is the frequency of the radiation emitted in the transition from the state or orbit of effective quantum number n^* to that of effective quantum number n'^*. K is the so-called Rydberg constant $=1.09\times10^5$cm.$^{-1}$. The two quantities $K/(n^*)^2$ and $K/(n'^*)^2$ are called "spectral terms," and it is from them that the experimental values of the effective quantum numbers are obtained.

If we were to assume at first that the 2_1 orbits lay completely inside the 2_2 orbits since the mutual screening of the electrons in the latter is taken as 1.5 units, this would mean that these electrons would move in a field corresponding to an effective nuclear charge of $11-2-4-1.5=3.5$ units. The corresponding orbits would thus have a radius of $4a_0/3.5=1.15a_0$. This radius, however, is, if at all, only a little larger than the aphelial distances of the 2_1 orbits calculated by neglecting the presence of the 2_2 orbits. In fact a rough calculation on this assumption gives a value of the aphelial distance for the 2_1 orbits about equal to $1.0a_0$. In such a case it is clear that the polarization induced in the inner configuration by the 2_2 orbit electrons and neglected in our estimation would bring the latter into the region of the 2_1 orbits. This, however, changes completely the basis for our considerations, as the only equilibrium configuration of the 2_2 orbits in this region will practically coincide with that calculated by neglect of the presence of the 2_1 orbits. This would mean that the 2_2 orbits would be in a field corresponding to an effective nuclear charge of $11-2-1.5=7.5$ units, which gives a radius of $0.53a_0$. The 2_1 electrons will during the greater part of each revolution move in a field, which apart from their mutual screening will correspond to a nuclear charge of 5 units. Since a 2_1 orbit under such a field would have a perihelion distance of about $0.11a_0$, we see that the actual 2_1 orbits even with neglect of any polarizing effect will penetrate into the region of the 2_2 orbits. A closer consideration shows, indeed, that they actually penetrate into the region of the 1_1 orbits and approach the nucleus to a distance of very nearly $0.05a_0$. The time during which the 2_1 electrons remain in the latter region is, however, so short that they will influence the 1_1 and 2_2 electrons only very slightly. Even within the region of the 2_2 orbits, the 2_1 electrons, as a rough calculation shows, will remain each only about $T/10$ where T is the time of one revolution; their screening effect on the 2_2 orbits may therefore be estimated to a first approximation as 0.5 unit, making the effective nuclear charge to which the 2_2 orbits are exposed equal to 7 units instead of 7.5 units. This will give a shell radius slightly greater than that indicated above and equal to $0.57a_0$. It is to be noted that this assumption of the

screening of the penetrating 2_1 electrons is only for a first approximation. As soon as an approximation to the correctly quantized 2_1 orbit is obtained, it is easy to determine exactly how much time the electron in such an orbit will spend inside a shell of radius 0.57 a_0. Any necessary correction will then be made in the radius of the shell in making the next approximation to the 2_1 orbit. It may be mentioned here, however, that such corrections will in general be very small.

To come to the specific discussion of the 2_1 orbits, our assumption leads to a mutual screening effect of 3 units for the electron in the extreme outer part of such an orbit, and hence to an effective nuclear charge of 2 units for this part. In accordance with the method described in Section I, we have next to assume what appears, by previous rough calculation, to be a reasonable value of the effective quantum number. We begin with $n^*=1.24$. The total energy of the electron in the elliptic orbit corresponding to this choice is then fixed at $E=-1.30e^2/a_0$, from the equation $E=-N^{*2}e^2/2a_0n^{*2}$, which is derived by a simple transformation from Equation (7). The dimensions of the orbits are also fixed according to (3) and (4). Spherical distributions are now found for each of the four electrons in accordance with our assumption. The potential energy V of any single 2_1 electron at a distance r from the nucleus is then calculated by finding the potential due to the distribution of the other three 2_1 electrons, and that due to the inner groups of 1_1 and 2_2 orbits, already previously disposed of (using 0.57 for the radius of the 2_2 shell). In this way Q is determined and may be plotted as a function of r. Graphical integration then leads to a value of $n-k=1.06$ instead of 1.00. It is evident that the orbit as chosen is too large. Another trial is made, starting with $n=1.21$ and hence $E=-1.37e^2/a_0$. This yields $n-k=0.985$ or fairly near to correct quantization. This corresponds to an aphelial distance of $1.14a_0$ for the 2_1 orbit.

The following table (Table 3) gives the values of V as a function of r for this case:

TABLE 3

r	V
.1	77.65
.2	35.59
.3	18.90
.4	11.56
.5	7.17
.6	4.70
.8	2.99
1.0	2.10
1.14	1.75

In this and all following tables r is expressed in multiples of a_0 and V in multiples of e^2/a_0. The values of V are, of course, all negative, though for the sake of simplicity their absolute values only are recorded in the table.

Since we now know the field rather accurately from point to point, we can at once compute the real orbit point by point and determine the distribution of the charge over this orbit instead of the previous approximate elliptic orbit. At the same time we can determine what portion of the time the 2_1 electrons remain inside the 2_2 shell. In the case above, it is found that each one spends $0.18T$ inside instead of $0.1T$ as was assumed in the first approximation. This will put the 2_2 shell at $0.59a_0$. Using the new distribution and the new value of the 2_2 shell, $n-k$ is again calculated, with the result that correct quantization within the limits of accuracy of the graphical integration is attained, and the course of V as a function of r duplicates that given in Table 3, thus proving that the orbit is not only quantized but consistent in the sense explained in Section I. We may therefore set the energy of the 2_1 orbit at $-1.37e^2/a_0$ and its $r_{max}=1.15a_0$.

In considering next the 3_1 orbit we meet a sharp contrast to the case of lithium, for here the outer electron will penetrate into the inner region. For if, indeed, it moved wholly outside the rest of the atom under a field corresponding to an effective nuclear charge of 1 unit, its perihelion distance would still be $0.51a_0$. Table 4 gives the course of the potential energy function V for the 3_1 orbit. It will be noted that the tabulation extends

only to $1.2a_0$, since for values of $r > 1.15a_0$, by our assumption, V is simply e^2/r. This procedure will likewise be adopted, for the sake of conserving space, for all following tables where similar circumstances occur.

TABLE 4

r	V
.05	182.64
.10	76.22
.20	32.47
.30	17.89
.40	10.60
.50	6.37
.60	3.73
.80	2.04
1.00	1.17
1.20	.83

In calculating the aphelial distance for the actual orbit, the problem is simpler than in the other cases, since there is but one electron. We estimate first $n^* = 2$; the outer field being known to be 1 unit, we are able at once to calculate the Q function for the 3_1 electron. Integration yields $n - k = 1.88$ instead of 2. A second approximation, using $n^* = 2.10$, yields exact quantization, and therefore this is the value of the effective quantum number of the 3_1 orbit on the basis of our assumption. The experimental value for this quantity is 1.63, emphasizing again the inability of the present method to attain to correct absolute results for the effective quantum numbers. As in the case of lithium, however, we may refer to the importance of a consideration of the polarization effect in removing this discrepancy (see Section X).

IV. **The Potassium Atom** ($N = 19$). We shall now indicate briefly the application of the method outlined to the case of the orbits in the potassium atom.

As in the case of sodium we are able on our assumption to fix at once the 1_1 orbits in potassium. Assuming as before a mutual screening effect of 0.5 unit we obtain for the radius of these orbits, $a_0/18.5 = 0.054a_0$.

In considering next the group of 2-quantum orbits, we have first to fix the 2_2 orbits. In accordance with our assumption, the

mutual screening effect of the 2_2 orbits will be 1.5 units. Estimating for a first approximation that each of the 2_1 electrons penetrates inside the 2_2 orbits for approximately 1/10 of its period of revolution, we shall have an additional screening of 0.4 unit. (We here neglect the small penetration effect of the outer groups, the 3-quantum orbits, etc., in conformity with our general assumption.) This will make the effective charge on the 2_2 electrons $17-1.5-.4=15.1$ units and the radius will be $a=4a_0/15.1=0.265a_0$. As in the corresponding case in sodium, the above is only a first approximation to be corrected as soon as the approximate form of the 2_1 orbit is obtained.

The next step is the fixation of the aphelia of the 2_1 orbits. The method employed is identical with that used in the case of the 2_1 orbits in sodium. First a value is assumed for the energy of the orbit which seems likely to give the desired fit, and the usual distribution is made. On the quantum integral coming out to be very close to unity, the complete orbit is calculated from point to point by means of the field values obtained in the first approximation. Quantization again resulting we finally have the orbit fixed on our assumption. In this way it is found that the electron in the actual 2_1 orbit penetrates within the 2_2 orbits 0.19 of its period of revolution, and hence the size of the 2_2 shell stands corrected at $4a_0/14.74=0.271a_0$. The final value of the 2_1 orbit comes out to be that with energy $E=-18e^2/a_0$ and $r_{max}=0.50a_0$. For purpose of reference we insert below the table of the values of V for the 2_1 orbit (Table 5).

TABLE 5

r	V
.05	316.85
.10	145.72
.20	61.82
.30	35.49
.40	25.57
.50	20.00

We next investigate the 3_1 and 3_2 orbits. The first question arising here is: which of the two groups has its aphelia outside the other? We might at first expect that since the 3_1 orbits are more eccentric than the 3_2 orbits they would have greater aphelia

than the 3_2 orbits. But if we assume in conformity with this expectation values for the energy and aphelia of these two subgroups of orbits, even very rough calculations of the resulting quantum integrals show at once that the value for $n-k$ on this assumption will be entirely too large for the 3_1 orbit and entirely too small for the 3_2 orbit. Since the inner field on the 3_1 and 3_2 orbits depends more on the nuclear charge and the screening of the 2-quantum orbits than on the 3_1 and 3_2 orbits themselves, it is clear that proper quantization of the 3_1 and 3_2 orbits can only be attained by taking a larger aphelial distance for the latter and a smaller aphelial distance for the former. If this is done it is found that correct quantization is very nearly attained for a 3_1 orbit of energy $E=-1.23e^2/a_0$ and $r_{max}=1.80a_0$ and a 3_2 orbit of energy $E=-0.49e^2/a_0$ and $r_{max}=2.30a_0$. The fields on the 3_1 and 3_2 orbits now being found, the more exact orbits are computed point by point as in the case of the 2_1 orbits discussed above. New distributions for the 3_1 and 3_2 groups are then obtained and a second approximation to the quantum integrals is evaluated. This process of approximations is continued until there are obtained orbits not only correctly quantized but so adjusted that when the electrons are distributed in them according to our assumption the values of the field suffer no change. The following table (Table 6) gives the energies and field values for the orbits finally obtained:

TABLE 6

Orbit	3_1	3_2
E	1.30	.494
r	V	
.05	304.1	304.1
0.1	135.53	135.01
0.2	52.74	52.51
0.3	27.50	27.30
0.4	18.01	17.81
0.5	12.95	12.76
0.6	10.11	9.91
0.8	6.56	6.37
1.0	4.51	4.34
1.2	3.18	3.04
1.5	1.99	1.87
1.7	1.49	1.41
2.0		1.06
2.2		.91

The values of the aphelia corresponding to the correctly quantized 3_1 and 3_2 orbits are for 3_1, $r_{max}=1.70\ a_0$; for 3_2, $r_{max}=2.20a_0$. The dimension of the potassium ion is thus fixed by our method at $2.20a_0$.

The following table (Table 7) gives the values of the field on the 4_1 electron on our assumption.

Using these values, we get correct quantization for a 4_1 orbit with energy $E=-0.108e^2/a_0$ corresponding to a value of n^* of 2.15. It is seen that as in sodium this value is essentially higher than that ($n^*=1.80$) obtained from the spectrum. At the same time it will be seen that the result presents an interesting illustration of the slow increase in the effective quantum number of the outer orbits as we pass from one alkali to another in spite of the increase of one unit in the principal quantum number.

TABLE 7

r	V
0.1	134.19
0.2	51.69
0.3	26.58
0.4	17.15
0.5	12.12
0.6	9.30
0.8	5.78
1.0	3.77
1.5	1.35
2.0	0.58
2.5	0.400
3.0	0.333
3.5	0.286
4.0	0.25
4.5	0.220
5.0	0.200
6.0	0.167
7.0	0.143
8.0	0.125

V. **The Rubidium Atom** ($N=37$). We are able to fix at once the 1_1 orbits on our assumption, their radius proving to be $a_0/36.5=.027a_0$.

When we consider next the 2-quantum orbits, if we estimate for a first approximation that each of the 2_1 electrons penetrates inside the 2_2 orbits for 0.2 of its period of revolution, we fix the effective charge on the 2_2 electrons at $35-1.5-0.8=32.7$ units.

The radius will therefore be $4a_0/32.7 = 0.122a_0$. The 2_1 orbits are computed as in the case of potassium and sodium, and it is found that each of the 2_1 electrons actually penetrates the 2_2 orbits for 0.24 of its period. Hence the 2_2 radius must be corrected to $0.123a_0$. The correct 2_1 orbit on our assumption proves to be that with energy $E = -117.5e^2/a_0$ and $r_{max} = 0.23a_0$. The following table (Table 8) gives the course of V for this group of orbits:

TABLE 8

r	V
.02	1714.90
.04	817.70
.05	643.90
.10	297.40
.20	140.30

Coming next to a consideration of the 3-quantum orbits, we meet with a new feature, *viz.*, the fixation of the 3_3 orbits. The problem arising here is: will the 3_3 orbits lie inside or outside the aphelia of the 3_1 and 3_2 orbits? To answer this question, let us first assume that the 3_3 shell lies wholly *outside* the 3_1 and 3_2 orbits. A short calculation then shows that on this assumption the radius of the 3_3 orbits will be very nearly $0.75a_0$. But we next find that if we compute approximately the 3_1 and 3_2 orbits on the assumption that the 3_3 orbits lie *inside* their aphelia, we get a result differing very little from the value above, leading us, therefore, to expect a much larger value for the 3_1 and 3_2 aphelia if we had assumed the 3_3 orbits to be *outside* where they would have a smaller screening effect. The conclusion is clear that the 3_3 orbits on our assumption can be fixed only inside the aphelia of the 3_1 and 3_2 orbits. We are, therefore, led to take for a first approximation to the effective charge on the 3_3 orbits $37-10-2.5-2.4=22.1$ units where the 10 charges represent the total screening of the 1- and 2-quantum groups, the 2.5 charges the mutual screening of the 3_3 orbits, and the 2.4 the screening due to the penetration of the 3_1 and 3_2 orbits (assuming for a first approximation that each of these spends 0.2 of its period inside the 3_3 shell). This makes the radius of the 3_3 orbits equal to $9/22.1a_0 = 0.407a_0$.

The fixation of the 3_1 and 3_2 orbits involves no new difficulties

and by the method fully discussed in connection with the 3-quantum orbits in potassium, we finally arrive at the following corrected values for the 3-quantum orbits in rubidium; for the 3_1 orbit, $E=-13.67e^2/a_0$, $r_{max}=0.69a_0$; for the 3_2 orbit, $E=-10.11e^2/a_0$, $r_{max}=0.71a_0$; for the 3_3 orbit, $r_{max}=0.414a_0$. Table 9 gives the course of V for the 3_1 and 3_2 orbits.

TABLE 9

Orbit	3_1	3_2
E	13.67	10.11
r	V	
.02	1638.	1637.
.03	1044.	1042.
.05	589.70	589.
.10	251.50	251.38
.20	98.58	98.46
.30	53.92	53.87
.40	32.91	32.87
.50	23.25	23.22
.60	17.68	17.67
.70		14.35

In the fixation of the 4-quantum orbits we encounter no new difficulties. The following table (Table 10) gives the energies and field values for the orbits finally obtained:

TABLE 10

Orbit	4_2	4_2
E	1.15	0.463
r	V	
0.015	2200.	
0.02	1607.8	
0.03	1022.1	
0.05	574.6	
0.10	241.37	241.21
0.20	90.94	90.77
0.3	47.11	46.94
0.4	26.69	26.52
0.5	17.36	17.20
0.6	12.08	11.91
0.8	7.00	7.19
1.0	5.23	5.06
1.2	3.87	3.73
1.5	2.52	2.41
1.8	1.70	1.61
2.0	1.34	1.26
2.2		1.05
2.5		0.84
2.7		0.74

The corresponding aphelial distances are for the 4_1 orbit, $2.04a_0$; for the 4_2 orbit, $2.74a_0$.

The dimension of the rubidium ion is thus fixed on our assumption at $2.74a_0$.

The fixation of the 5_1 orbit is similar in every way to the fixation of the 4_1 orbit in potassium and the 3_1 orbit in sodium. It will not be necessary therefore to go into details, but merely to state that we get correct quantization for a 5_1 orbit with energy $E=-0.103e^2/a_0$ corresponding to an effective quantum number of 2.20.

Table 11 gives V as a function of r for the 5_1 orbit.

TABLE 11

r	V
.02	1604.1
.03	1019.4
.05	572.8
.10	240.2
.20	90.04
.30	46.32
.40	25.95
.50	16.65
.60	11.39
.80	6.69
1.00	4.60
1.20	3.27
1.50	1.96
1.80	1.18
2.00	.84
2.50	.46

VI. **The Caesium Atom** ($N=55$). The 1_1 orbits are fixed at once on our assumption with a radius of $a_0/54.5=0.018a_0$.

In considering next the 2-quantum orbits, the effective charge on the 2_2 orbits is taken to a first approximation as $53-1.5-0.8=50.7$ units (assuming that each of the 2_1 electrons penetrates inside the 2_2 orbits for 0.2 of its period). This fixes the radius of the 2_2 orbits at $0.079a_0$. Proceeding in the usual way, we arrive at the correctly quantized 2_1 orbit with $r_{max}=0.163a_0$ and energy $E=-264.5e^2/a_0$. From the overwhelming influence of the nucleus, it is clear that this orbit is practically a Keplerian orbit with effective quantum number 2 equal to the principal quantum

number. Therefore a table of the course of the V function is omitted.

The fixation of the 3-quantum orbits is carried out as in the case of rubidium; hence it will not be discussed in detail. The final values are as follows: 3_1 orbit, $E=-69.9e^2/a_0$, $r_{max}=0.39a_0$; 3_2 orbit, $E=-57.5e^2/a_0$, $r_{max}=0.40a_0$; 3_3 orbit, radius $=0.224a_0$.

The following table (Table 12) gives the values of V for these orbits:

TABLE 12

Orbit	3_1	3_2
E	69.9	57.5
r	V	
.05	896.	896.
.10	391.	390.4
.20	161.8	161.3
.30	98.1	97.9
.40	70.3	70.2

The only new problem arising in connection with the 4-quantum orbits is that of the 4_3 orbits. The first approximation here is conducted more easily if we note that, unlike the 4_1 and 4_2 orbits, the 4_3 orbits do not penetrate into the region of the 1- and 2-quantum orbits (see Bohr, loc. cit., page 262). The following table gives the field values for the 4-quantum orbits at various distances from the nucleus:

TABLE 13

Orbit	4_1	4_2	4_3
E	10.41	8.3	5.28
r		V	
0.1	355.70	355.64	355.25
0.2	132.02	131.96	131.81
0.3	70.98	70.92	70.77
0.4	44.81	44.76	44.60
0.5	32.45	32.40	32.27
0.6	24.39	24.36	24.24
0.8	14.88	14.86	14.78
1.0	10.29	10.29	10.25
1.1			9.14

The corresponding aphelial distances are: for 4_1 orbit, $r_{max}=0.98a_0$; for 4_2 orbit, $r_{max}=1.00a_0$; for 4_3 orbit, $r_{max}=1.16a_0$.

It is unnecessary to enter into the details of the quantization of the 5-quantum orbits. The final values are: for 5_1, $E = -0.925e^2/a_0$, $r_{max} = 2.50a_0$; for 5_2 orbit, $E = 0.415e^2/a_0$, $r_{max} = 3.40a_0$. Table 14 gives the course of V for the 5_1 and 5_2 orbits.

TABLE 14

Orbit	5_1	5_2
E	0.925	0.415
r	V	
.10	346.76	346.67
.20	125.57	125.45
.30	65.38	65.25
.40	39.65	39.52
.50	27.54	27.42
.60	19.63	19.51
.80	10.42	10.29
1.00	6.06	5.94
1.20	4.33	4.21
1.50	2.99	2.88
1.80	2.12	2.01
2.00	1.76	1.61
2.50	1.04	.98
3.00		.72

It is thus seen that on our assumption the dimension of the caesium ion is $3.40a_0$.

A discussion of the 6_1 orbit shows that the energy of this orbit when correctly quantized is given by $E = -0.0945e^2/a_0$, corresponding to an effective quantum number 2.31. Table 15 gives V for the 6_1 electron orbit.

TABLE 15

r	V
.10	345.71
.20	124.84
.30	64.75
.40	39.08
.50	26.99
.60	19.11
.80	9.91
1.00	5.58
1.20	3.87
1.50	2.50
1.80	1.65
2.00	1.31
2.50	0.64
3.00	0.40
3.50	0.29

VII. Summary of Sections I-VI with Discussion of Results. The object of this paper has been to derive on a certain simple assumption the magnitude of the aphelial distances of the various types of orbits in the Bohr models of the alkali metals. In the carrying out of this work we are also naturally led to values for the effective quantum numbers for the outermost orbits in this family. On the Bohr theory, as has been noted in the introduction, these effective quantum numbers are identical with the square roots of the denominators of the first spectral terms in the sharp series of the arc spectra of these elements (see Bohr, loc. cit. page 264; see also Note 14 of this dissertation). The comparison between the results of the present paper and the spectroscopically determined values may be summarized in the following table. The values of n^* observed from the series terms are taken from Fowler[15].

TABLE 16

Element	n	n^* *Calc.*	n^* *Obs*
Lithium	2	2.00	1.59
Sodium	3	2.10	1.63
Potassium	4	2.15	1.77
Rubidium	5	2.20	1.81
Caesium	6	2.31	1.87

While there is no absolute agreement between the experimental values and those deduced in this paper, there must be noted the significant fact that while the principal quantum number n of the outermost orbit increases by unity as we pass from each alkali metal to that next succeeding, the effective quantum numbers, like those obtained from the series terms, increase very slowly, remaining continually in the vicinity of two. As has been pointed out by Bohr[16], there lies in this circumstance an instructive example of the explanation of the periodicity in the properties of the elements with increasing atomic number. In connection with the discrepancy between the observed values and the calculated values of n^*, we may recall that the deviation in the case of lithium may be largely explained by taking into account the polarization of the inner configuration by the outermost electron.

[15] Report on Series in Line Spectra, Phys. Soc. of London, 1922.
[16] Loc. cit. page 265.

As has already been mentioned, a calculation of the induced polarization and its effect on the value of n^* has been carried out for lithium and sodium in Section X. It is there found that the discrepancy between our values of n^* and the observed values may be removed completely.

More important and significant than the effective quantum numbers, however, are the results obtained for the dimensions of the various types of orbits in the normal alkali atoms. These values are summarized briefly in the following table:

TABLE 17

Element	1_1	2_1	2_2	3_1	3_2	3_3
Lithium	.40	7.46				
Sodium	.095	1.15	.59	8.34		
Potassium	.054	0.50	.27	1.70	2.20	
Rubidium	.027	0.23	.12	0.69	0.71	.41
Caesium	.018	0.16	.079	0.39	0.40	.22

Element	4_1	4_2	4_3	5_1	5_2	6_1
Potassium	8.77					
Rubidium	2.04	2.74		9.26		
Caesium	0.98	1.00	1.16	2.59	3.40	10.21

The values here given are the aphelial distances or r_{max} of the types of orbits noted in the top row, and are all multiples of a_0. From the above discussion on the effective quantum numbers it is clear that the values for the r_{max} of the outermost orbits, *i.e.*, the dimensions of the atoms in the normal state, are much too large. Even here, however, it is significant to note, they all remain in the region between $7.5a_0$ and $10.00a_0$. The values for the inner orbits may be looked upon with more confidence, though naturally they must be judged wholly on the fundamental assumption used in obtaining them. It will be of interest at this point to compare the results of the above-mentioned work of Hartree (see Section I) with those just tabulated. Hartree calculates r_{max} for the orbits in sodium and potassium only (in the alkali family), and his values are given in the following table (Table 18), the arrangement of which parallels that of Table 17:

TABLE 18

Element	1_1	2_1	2_2	3_1	3_2
Sodium	0.09	1.20	0.75		
Potassium	0.055	0.50	0.30	1.8	1.75

When we consider the fundamental difference in method, the agreement between the two tables is very satisfactory. The essential discrepancy appears in the 3_2 orbit in potassium. In the discussion of his results, Hartree admits uncertainty about just this value. From the standpoint of the present research there seems no doubt that the 3_2 orbits have aphelia lying outside those of the 3_1 orbits. Indeed, our results in general indicate the general rule that for any two types of orbits n_k and $n_{k'}$ $(n \neq k \neq k')$ if $k > k'$ the orbit n_k will have a greater aphelion than $n_{k'}$. This point will bear further investigation since it should enter into a consideration of the *spark* spectra of the alkali metals where we encounter transitions between orbits of the kind mentioned.

A comparison of the dimensions of the alkali *ions* as calculated on our assumption with those calculated in other ways will be instructive. The following table contains the values of the radii of the positive ions of the alkali metals as computed by Grimm[17] on the basis of crystal structure, and by Landé[18] also on crystal structure data, together with the values obtained in the present research:

TABLE 19

Element	*Grimm*	*Landé*	*Present Research*
Sodium	0.98	2.05	1.15
Potassium	1.51	2.69	2.20
Rubidium	1.73	2.84	2.74
Caesium	2.01	3.40	3.40

As above the values of the ionic radii are in multiples of a_0.

[17] H. Grimm, ZS. f. Phys. Chem. 5, 353, 1921.
[18] A. Landé, ZS. f. Phys. 1, 191, 1920.

The better to facilitate comparison, the three sets of values are plotted to the same scale in the following graph (Fig. 3):

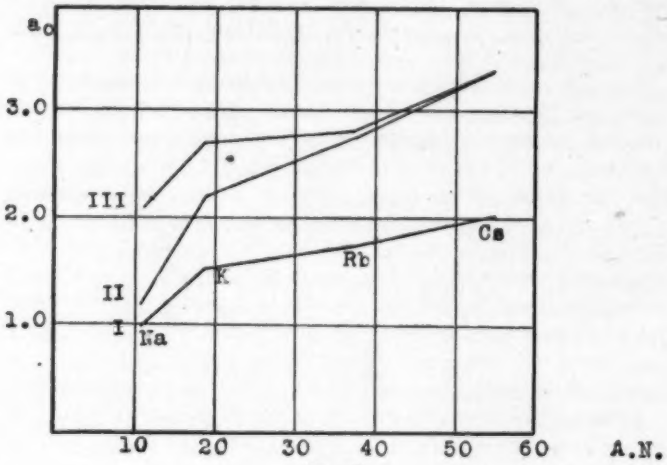


Fig. 3

In the figure the abscissæ are atomic numbers, the ordinates are ionic radii in multiples of a_0. Curve I refers to Grimm's values; Curve II to those of the present paper; Curve III to Landé's values. It is to be observed that the absolute agreement between any two of the three sets is poor, save that Landé's values for rubidium and caesium agree rather well with those of the present paper, while Grimm's value for sodium agrees fairly well with our value. More important, however, is the agreement in trend between the three curves. If we denote the ionic radii by R_{Na}, R_K, etc., then Grimm deduced from his work the following inequality law, holding generally for the ionic radii in any given family of elements (but here applied only to the case of the alkali ions):

$$R_K - R_{Na} > R_{Cs} - R_{Rb} > R_{Rb} - R_K.$$

This inequality is readily observable in the figure as well for the results of the present paper as for those of Grimm and Landé.

This fundamental agreement is of great interest because of the wide applicability of the Grimm law. As is well known, this inequality is observed not only in the measurements of the grating constants of many crystals (whence the law in its above form was first enunciated), but also for many other physical properties of families of elements, notably, the melting point, boiling point, atomic refraction, atomic dispersion, atomic susceptibility, volume in the liquid state, etc.

In making any detailed comparison of the results of Grimm and Landé with those of the present research, it must be constantly borne in mind, of course, that the quantity representing the size of the alkali ion from the point of view of this work is the aphelial distance of the greatest orbit present in the unexcited ion, whereas in the results of the above-mentioned authors this quantity is taken to be some function of the grating space in crystals of alkali salts. Thus Landé assumes that the grating space, *i.e.*, the distance between two oppositely charged neighboring ions in a crystal, is equal to the sum of the radii of the spheres of action of the two ions; hence from an empirical knowledge of these grating space constants coupled with an estimate of the radius of the action sphere of one of the ions in question (in Landé's paper, iodine in the salt, lithium iodide), the action spheres of the various ions may be computed. The radii of these spheres are then taken as a measure of the radii of the ions[19]. Landé's hypothesis is not, of course, exact, assuming as it does that the ions are so closely packed in a crystal that their action spheres are in contact. From a comparison of his values of the radii with ours, it would seem that for the lighter alkalies the ions are not close packed, while the good agreement for the case of the heavier alkalies indicates that the latter ions are close packed.

Coming next to a consideration of Grimm's method we note that he takes as a measure of the radius of the ion, at least in the case of deformable ions, a quantity measuring the intensity of the repulsion of two neighboring ions of the same sign. A comparison between his values and ours indicates that for the heavier

[19] It should be noted here that the values attributed to Landé in Table 19 are not those to be found in his paper above cited, but are the corrected values calculated by Davey (Phys. Rev. September, 1923, p. 211) from newer and more accurate values of the grating constants.

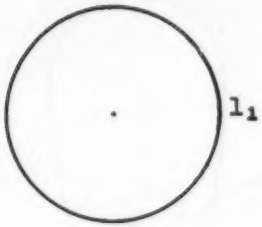
alkalies this repulsion is abnormally small and no longer gives a true measure of the dimensions of the outer orbits in the ion. This is what we might expect since for ions with larger orbits the latter would overlap those of neighboring similar ions and by their interplay produce a weakening of the repulsion normally to be expected.

From this discussion it is evident that our results are not inconsistent with the points of view of Grimm and Landé.

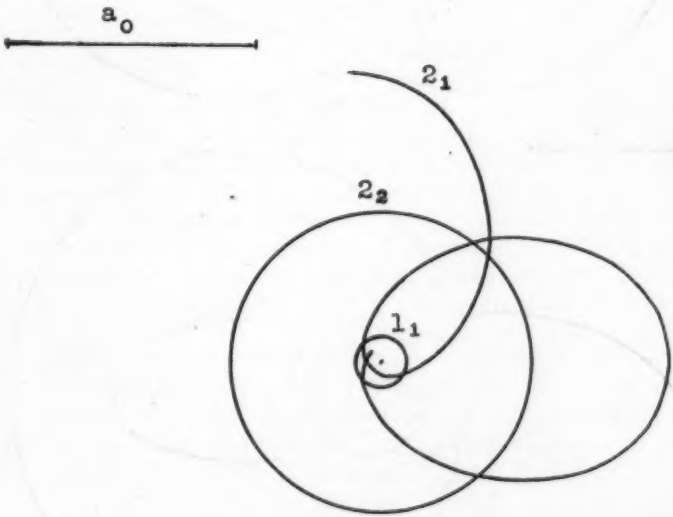
It lies far outside the scope of the present research to analyze the great mass of work that has been done in recent years on the estimation of atomic and ionic radii from various points of view. Good surveys of this field are to be found in the articles of Davey[20] and Lorenz[21], who go fairly exhaustively into the literature of the subject. In particular, Lorenz calls attention to the above-noted distinction between the ionic radius as obtained on the close packing idea and that based on the dimensions of the electron orbits in the ion. In this connection it is of importance to note that in a more recent paper Landé[22] has attacked the ionic radius problem from the standpoint of the application of the quantum theory to the crystal grating, and considers the outer eight electrons in the sodium ion as moving in elliptic orbits arranged in cubical symmetry (*i.e.*, revolving about the corners of a cube). In this way he deduces for the dimension of the sodium ion the value $1.14a_0$, in almost exact agreement with the value obtained in the present research (*i.e.*, $1.15a_0$). So far as the knowledge of the author goes, Landé has not applied his new method to the other alkali metals, thus unfortunately precluding further comparison.

VIII. **Diagrams of the Orbits.** In this section are presented drawings of the various orbits in the ions of the alkali metals, obtained by the use of Equation (26). The scale used is noted on each page, and for the sake of simplicity of presentation one only of each type of orbit is drawn.

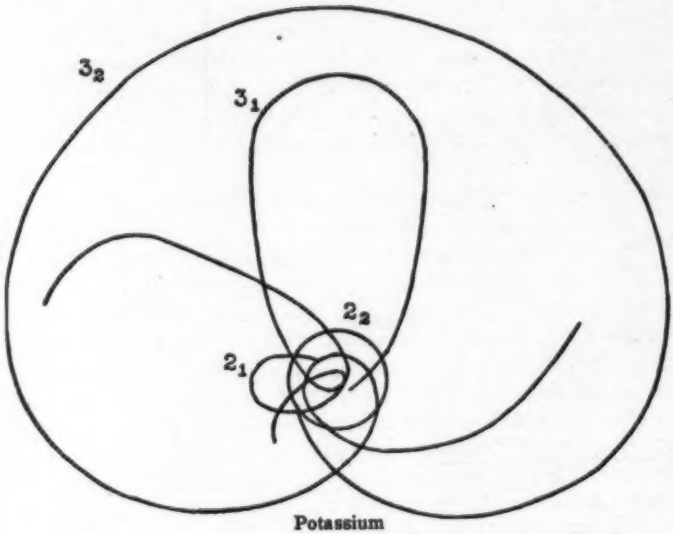
[20] W. P. Davey, Phys. Rev. September, 1923, p. 211.
[21] R. Lorenz, ZS. f. Phys., 6, 271, 1921.
[22] A. Landé, ZS. f. Phys., 2, 380, 1920.



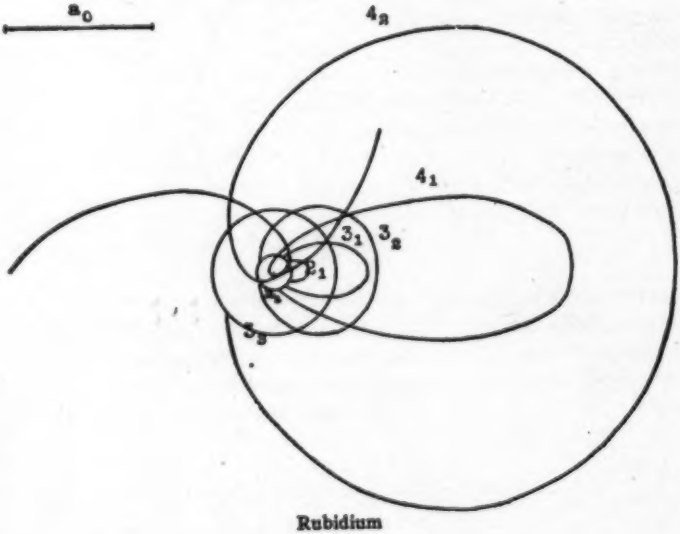
Lithium



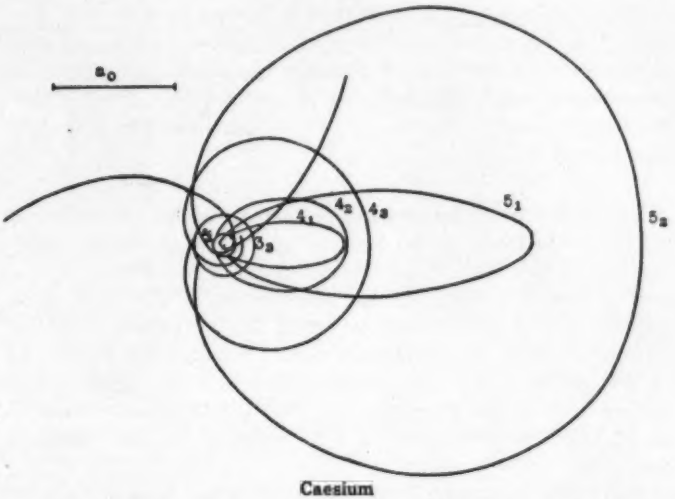
Sodium



Potassium



Rubidium



Caesium

The following comment may be of help in the interpretation of the preceding sketches. To simplify the figures the circle representing the 1_1 orbits has been omitted for al atoms subsequent to sodium. The same is true of the 2_1, 3_1 and 3_2 orbits in caesium. To prevent needless complexity no attempt has been made to indicate the penetration of the inner groups of orbits, the perihelia of the outer orbits only being indicated. In the case of the latter the inner loops are extended to show the extent of rotation of the orbit.

The following table (Table 20) may be of interest in connection with the rotation of the orbits. This gives the values of G, the half-angle of rotation of the major axis of each orbit for one revolution (where by major axis is here meant the radius vector containing the aphelion). The values of G are given in radians.

TABLE 20

Element	2_1	3_1	3_2	4_1	4_2	4_3	5_1	5_2
Sodium	0.81							
Potassium	0.43	0.66	3.48					
Rubidium	0.23	0.47	1.32	0.67	2.12			
Caesium	0.10	0.19	0.46	0.55	0.84	2.31	0.57	2.10

For any given type of orbit it is of interest to note the progressive decrease in G from the light alkalies to the heavier, indicating for the inner groups of orbits the steady progression to orbits more nearly Keplerian in character. This circumstance has already been alluded to in the Introduction (see page 192) and Section III.

IX. **Fixed Phase Relations in the Sodium Atom.** In Section I, reference was made to the question of the phase relations existing among the electrons in a given group of non-circular orbits, and a simple hypothesis for accounting for this was mentioned. It is here proposed to utilize this hypothesis to calculate the 2_1 orbits in the sodium atom and hence the dimensions of the sodium ion. Our fundamental assumption here is that the electrons in the 2_1 orbits attain their perihelia successively at equal intervals of time, *i.e.*, $T/4$, where T is the period of revolution of the electron.

We begin by noting that the radii of the 1_1 orbits will be unaffected by this change in assumption, *i.e.*, the value will remain at $0.095\ a_0$ (see Section III). Moreover, for the first approximation, we assume that the 2_2 orbits have the same radii as previously computed in Section III, *i.e.*, $0.59a_0$. In calculating the 2_1 orbits, our main procedure is much the same as before, that is, to assume as a first approximation a Keplerian orbit, make the spherical distribution, calculate the field function V and evaluate the integral (13), conducting successive approximations until a fit is obtained. The essential difference appears in the computation of V. We have here to account for the relative distances from the nucleus of the four electrons during each revolution. We shall assume that each electron, as far as its influence at any moment on the other electrons is concerned, is replaced by the uniformly charged spherical surface through which it is passing at that moment. If we keep this idea and Fig. 2 in mind, it is easy to construct the following table (Table 21) of screening potentials from which V may be found as a function of r[23]:

[23] The assumption of a central field is still retained.

TABLE 21

Interval			V_s = *Screening Potential*
r_{max}	to	$r_{T/16}$	$3e^2/r$
$r_{T/16}$	to	$r_{2T/16}$	$3e^2/r$
$r_{2T/16}$	to	$r_{3T/16}$	$2e^2/r + e^2/r_3$
$r_{3T/16}$	to	$r_{4T/16}$	$2e^2/r + e^2/r_1$
$r_{4T/16}$	to	$r_{5T/16}$	$e^2/r + e^2/r_1 + e^2/r_4$
$r_{5T/16}$	to	$r_{6T/16}$	$e^2/r + e^2/r_2 + e^2/r_5$
$r_{6T/16}$	to	$r_{7T/16}$	$e^2/r_2 + e^2/r_3 + e^2/r_6$
$r_{7T/16}$	to	r_{min}	$e^2/r_1 + e^2/r_4 + e^2/r_5$

The table gives the values of V_s, the mutual screening potential of the 2_1 electrons for the intervals noted in the left-hand column. These intervals run from r_{max} to r_{min} through intermediate values of r corresponding to a time interval of $T/16$. In the expression for V_s we shall understand by r the particular value of the distance from the nucleus for which V is desired. By r_1, r_2, r_3, r_4, r_5 and r_6 we shall mean the corresponding values of r for the time intervals $(0, T/16)$, $(T/16, 2T/16)$, $(2T/16, 3T/16)$, $(3T/16, 4T/16)$, $(4T/16, 5T/16)$, $(5T/16, 6T/16)$, respectively. These values are obtained directly from a study of Figure 2 and our usual distribution method (see (27.)

Table 22 gives the course of V for the final consistent, quantized orbit obtained on the above assumptions. This orbit proves to be that with energy $E = -1.50e^2/a_0$ and $r_{max} = 1.01a_0$. The calculation of the complete orbit shows that each 2_1 electron remains

TABLE 22

r	V
.10	79.96
.20	35.10
.30	20.00
.40	12.20
.50	7.80
.60	5.00
.80	2.90
1.00	1.99

inside the 2_2 shell $0.17T$, and hence the effective nuclear charge on the 2_2 electrons is 6.82 units, placing their radii at $0.586a_0$, or not very different from the value obtained by the work of Section III. The aphelion of the 2_1 orbit calculated as above

comes out to be about 15 per cent smaller than that calculated in Section III (*i.e.*, $1.15a_0$). It is interesting to note that the value here obtained (*i.e.*, $1.01a_0$) is in good agreement with Grimm's value for the effective radius of the sodium ion (see Table 19).

X. **The Polarization Effect in Lithium and Sodium.** In the discussion of the lithium atom (Section II) reference was made to the importance attaching to a consideration of the polarization induced by the outermost electron on the inner residue of the atom. In obtaining the main results of this research (Sections I–VIII), this effect was neglected. It is proposed in this section to consider it quantitatively in connection with the calculation of the effective quantum numbers of the 2_1 and 3_1 orbits in the lithium and sodium atoms respectively.

Let us first examine the general question of the polarization of an inner charged shell by an outer electron, for this provides the simplest way of handling our problem. Thus we shall consider that all the inner electron groups of the atom in question have their charges, q units in number, distributed uniformly over a sphere of radius a about the nucleus. As the outer electron moves in toward the nucleus, it polarizes this shell, and we must find the expression for the potential energy V of the electron as a function of r, the distance from the nucleus. Using the well-known method of images[24] we arrive at the following expression for this quantity:

$$V = -\frac{Ne^2}{r} + \frac{qe^2}{r} - \frac{e^2a}{r^2-a^2} \tag{28}$$

for $r > a$.

In this expression N is the number of unit charges in the nucleus and q the number on the shell. The term $e^2a/(r^2-a^2)$ is thus a kind of correction term to the ordinary expression V used in our previous work. The procedure now is to seek a value of E which, when substituted with V above into the Equation (13), yields upon integration the appropriate integral value for $n-k$. From this value of E, n^* may be immediately calculated from Equation (7) (note that $W = -E$). In applying (28), one ques-

[24] See Clerk Maxwell, Treatise on Electricity and Magnetism, Third Edition, Vol. I, page 244.

tion arises at once, *viz.*, how far from the nucleus are we to consider the polarization as still effective. It will be seen that in (28) $V=\infty$ for $r=a$, *i.e.*, as the electron passes through the shell. Since the shell has no real physical significance save when the electron is at some distance outside the shell, it is clear that trying to integrate through the shell using (28) will yield a result which has little or no significance. We ought, therefore, to set some rational limit to the validity of (28). One somewhat arbitrary, though for the present sufficiently rational, scheme is to use (28) from r_{max} to that value of r, say r_1, for which the resultant surface density at the point on the shell nearest the electron has become zero. From r_1 to r_{min} we shall assume that the distribution of electricity on the shell remains unaltered and the same as its distribution when the electron is at r_1. When it is considered that the time spent by the electron in passing from r_1 to r_{min} and back to r_1 again is very short relative to its whole period, and is of the order of magnitude of the periods of revolution of the inner electrons whose effect is here replaced by the shell, it will be seen that our assumption is about the simplest rational one at present available that still permits the use of the shell idea. The value r_1 above mentioned is obtained as the root of the equation

$$\sigma=\frac{e}{4\pi a^2}\left(q-\frac{a(r+a)}{(r-a)^2}\right)=0 \tag{29}$$

where σ is the resultant surface density at the point of the shell nearest to the electron. It may then be shown that from r_1 to r_{min} the expression (28) for V becomes

$$V=-\frac{Ne^2}{r}+\frac{qe^2}{r}-\frac{e^2(d^2-1)}{2a}\int_0^{\pi}\frac{\sin\theta\,d\theta}{(d^2+1-2d\cos\theta)^{3/2}(c^2+1-2c\cos\theta)^{1/2}}$$

$$=-\frac{Ne^2}{r}+\frac{qe^2}{r}-\frac{e^2(d^2-1)}{2a}f(r) \tag{30}$$

for $a<r<r_1$, and

$$V=-\frac{Ne^2}{r}+\frac{qe^2}{a}-\frac{e^2(d^2-1)}{2a}f(r) \tag{31}$$

for $r_{min}<r<a$.

In these equations $c=r/a$ and $d=r_1/a$, while θ is an auxiliary variable (the angle at the nucleus between the radius vector and any radius of the shell), and

$$f(r)=-\frac{1}{b\sqrt{dc}}\left[\frac{\sqrt{b+(d+1)^2}}{d+1}-\frac{\sqrt{b+(d-1)^2}}{d-1}\right],$$

where $$b=\frac{(cd-1)(c-d)}{d}.$$

We shall now apply the above to the relatively simple case of lithium. For this atom the shell of the 1_1 orbits has a radius of $0.4a_0$ (see Section II). Hence in the nomenclature used immediately above, $a=0.4a_0$. Moreover, $N=3$ and $q=2$. Solution of Equation (29) under these conditions yields $d=2.28$ or $r_1=0.912a_0$. Calculation of V as indicated above then yields the following table (Table 23):

TABLE 23

r	V
.2	11.43
.3	6.62
.4	4.78
.6	2.83
.8	1.95
1.0	1.48
1.5	.85
2.0	.61
3.0	.38
4.0	.28
5.0	.22

Assuming now $n^*=1.56$, whereby $E=-0.205\varepsilon^2/a_0$, we find that (13) yields $n-k=1$. In other words, the effective quantum number of the 2_1 orbit in lithium under the above assumption is 1.56, differing from that obtained from the series term (1.59) by less than 2 per cent.

It will be observed that we have here neglected the small screening effect of the 2_1 electron on the 1_1 electrons due to its penetration. This effect is negligible, due to the very short time that the outer electron spends within the shell.

The treatment of the 3_1 orbit in the sodium atom by the method above described is more difficult because of the presence of the

2_1 orbits. The latter may be taken with the 2_2 orbits to form a common resultant shell when the 3_1 electron is at relatively large distances from the nucleus. When the 3_1 electron approaches to within a distance a_0 of the nucleus, they may, however, not be taken as a shell. In the present work this difficulty is avoided in the following manner. For $r > r_1$ we shall assume that the 2_1 and 2_2 electrons are replaced by a shell of radius $0.72a_0$, which is the average of the radius of the 2_2 shell (*i.e.*, $0.59a_0$) and the time mean radius of the 2_1 orbit as calculated in Section II (*i.e.*, $0.85a_0$). For $r > r_1$, then, the potential V will be calculated by (28) substituting $N=9$ (the 1_1 electrons are here combined with the nucleus) and $q=8$. The appropriate value of r_1 is obtained as the root of (29) and is equal to $1.13a_0$ ($d=1.57$). For $r_{min} < r < r_1$ we shall assume the screening of the 2_1 orbits to be that obtained in Section II, while the 2_2 shell only is polarized by the 3_1 electron. For this part we use (28) substituting $N=9$, but $q=4$ and introduce as an additional term the screening of the 2_1 orbits, save that for $r_{min} < r < 1.09a_0$ we must use the modified forms (30) and (31) as explained above in the case of lithium. Table 24 gives the resulting values of V for the 3_1 orbit.

TABLE 24

r	V
.1	77.24
.2	33.18
.3	18.64
.4	12.14
.6	5.48
.8	3.16
1.0	2.14
1.5	1.08
2.0	.71
3.0	.42
4.0	.30
5.0	.23

If we now assume $n^* = 1.63$, corresponding to an energy value of $E = -0.189e^2/a_0$, we find that evaluation of (13) yields $n-k=2$. We may, therefore, state that the effective quantum number for the 3_1 orbit in sodium under our present assumption is 1.63, which agrees exactly with the observed value from the spectral series term (see Section II). As in the case of lithium we may neglect

the screening due to the penetration of the 3_1 electron within the 2_2 shell. A comparison between Table 24 and Table 4 is very instructive as to the influence on V of the consideration of the polarization effect.

The results obtained in this section, while not pointing to a complete solution of the question of polarization, indicate at any rate what an important part it plays in the fixation of the outer electron orbits, and how by simple yet rational assumptions we can explain the discrepancy between the values of n^* tabulated in Section VII (Table 19) and the values observed from the spectral terms.

Acknowledgments. The writer desires at this place to express his great indebtedness to Prof. Niels Bohr of the University of Copenhagen for his encouragement and assistance during the early stages of the preceding work. He is also very grateful to Prof. H. B. Phillips of the Massachusetts Institute of Technology, the supervisor of the work, and Prof. Leigh Page of Yale University for valuable suggestions.

ON DYNAMIC STRESSES IN PSEUDO-CONTINUOUS MEDIA[1]

By Paul Heymans

Part I

Summary

After an introductory presentation of the known theory of torsional oscillatory stresses and strains in non-dissipative and dissipative discontinuous elastic media, disturbed initially, or disturbed by a stationary periodic force, the theory is extended in this paper to the case of a constant cyclical force and a periodic cyclical force acting on discontinuous elastic media and on so-called pseudo-continuous elastic media.

In conclusion, it can in general be seen by inspection of the primitive Lagrangian equations of motion (9) that in any mechanical system similar to those considered, be they treated as discontinuous or pseudo-continuous, when any of the constituent parts vary in potential or in kinetic energy in any of their degrees of freedom, oscillatory displacements and dynamic stresses are set up.

If the medium consists of a system of independent mass particles — or independent rigid bodies or systems — equations of the type (15) express the kinetic equations of motion of the different parts.

If the medium is pseudo-continuous, *i.e.*, with elastic connections between the mass particles, equations (20) express the states of motion and the corresponding states of stress. The frequencies of the free components, forming the transient strain and stress components, are then given by the roots of the determinantal equation (23) $\Delta(p)=0$. The frequencies of the forced or stationary components are equal to the frequencies of the impressed force components. The amplitudes of the transient components are determined, as in the classical theory for discontinuous elastic media, by the connections of the system and its configu-

[1] Part of this article is included in a paper on the Mathematical Theory of Dynamic Stresses in Rotating Gear Pinions presented by the author at the Spring Meeting, 1924, of the American Society of Mechanical Engineers.

ration when the disturbing forces are applied; the amplitudes of the steady state components depend upon the frequencies of the impressed forces and the natural frequencies of oscillation. It is also shown that the stress and strain components, entering into the determination of the dynamic state of stress and the dynamic state of strain throughout the medium, cannot in general be arbitrarily chosen. The three principal stresses and the three principal dilatations alone, when known vectorially, fulfill the double condition that the stress or strain variables chosen propagate with the same velocity and fully define the elastic state at each point.

The theory presented has found its application in furthering the solution of the problem of dynamic stresses in revolving mechanical parts loaded by constant or periodic forces[2]. Formulæ for the derivation of dynamic stresses, such as have been proposed and used, which omit to take into account the necessary elastic constants of the material, the internal friction and the relation of the frequencies of the free oscillating system and the frequencies of the cyclical disturbing forces — whether constant or periodic — are, in the light of the present theory, at least incomplete in their essentials.

Static and dynamic stresses in discontinuous media of independent mass-points

Let us consider an ideal disc (Fig. 1) rigidly clamped at its center and assume the disc to be formed by *independent* mass particles obeying individually Hooke's law of linear proportionality between the elastic force and the displacement from their position in the unstressed state.

(1) Let a tangential force F_1 be applied at point (1) and result in a statical stress distribution in the disc in accordance with the theory of torsion.

If the force F_1 is removed, d'Alembert's equation of dynamical equilibrium between the inertia force and the elastic force gives as differential equation of motion for any one of the points of the disc:

[2] See Paul Heymans, loc. cit.

$$I\frac{d^2\theta}{dt^2}=-\frac{G\theta}{l} \tag{1}$$

where $\frac{G}{l}\theta$ is the restoring couple called into play by the elastic forces, θ being the angle of displacement, G the modulus of torque,

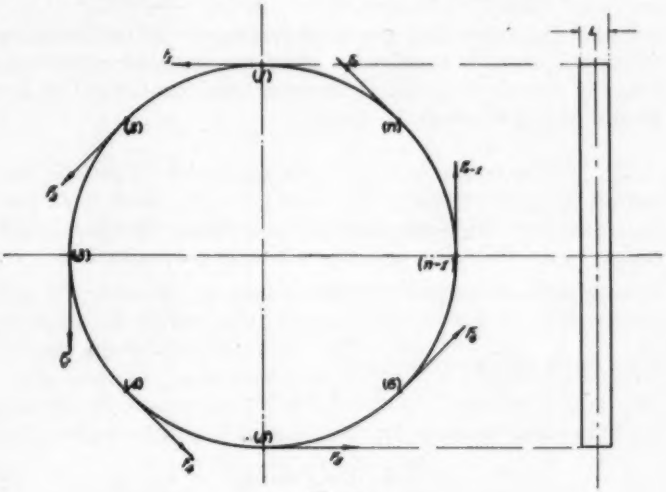


Fig. 1

l the longitudinal dimension of the disc, and I the moment of inertia about the axis of rotation.

The general solution of equation (1) gives as kinetic equation of motion:

$$\theta=A\cos\sqrt{\frac{G}{Il}}t+B\sin\sqrt{\frac{G}{Il}}t \tag{2}$$

where A and B are two arbitrary constants determined by the conditions of the motion at any given time at the point under consideration.

It follows from equation (2) that the different points will perform torsional oscillations around the longitudinal axis of the disc.

The period of the simple harmonic motion is:

$$T = 2\pi\sqrt{\frac{Il}{G}}.$$

The oscillation is solely due to the action of the elastic forces once the system has been initially disturbed. Such oscillations are called " free " oscillations.

It will be noted that whenever oscillatory deformations are shown to exist, the existence of the corresponding stresses, resulting directly from the stress-strain relations derived in the general theory of elasticity, follows.

(2) Let the tangential force F_1 be replaced by a periodic disturbing force $f(t)$ applied at the same point (1). Such a periodic disturbing force can be expressed, at least during its time of application, by an absolutely convergent Fourier's series such as:

$$f(t) = A_0 + A_1 \cos p_1 t + A_2 \cos p_2 t + \ldots\ldots$$
$$+ B_1 \sin p_1 t + B_2 \sin p_2 t + \ldots\ldots$$

If the series reduces to the term

$$f(t) = A \cos p\, t$$

the differential equation for this " forced " motion becomes:

$$I\frac{d^2\theta}{dt^2} + \frac{G}{l}\theta = A \cos p\, t. \tag{3}$$

The general solution of this equation is:

$$\theta = \frac{A/I}{p_0^2 - p^2}\cos p\, t + C \cos p_0 t + D \sin p_0 t \tag{4}$$

where p_0 is the circular frequency $\sqrt{\frac{G}{Il}}$ of the free oscillating system, whereas p is the circular frequency of the disturbing force.

If the function $f(t)$, expressing the periodic disturbing force, resolves into m harmonic terms, the kinetic equation of motion for any point becomes:

$$\theta = C \cos p_0 t + D \sin p_0 + \sum_{\nu}^{1,2,\ldots m} \frac{A_\nu/I}{p_0^2 - p_\nu^2} \cos p_\nu t \tag{5}$$

where $\nu = 1, 2, \ldots . m$.

If dissipative forces, proportional to the velocity, are introduced into equation (3), the differential equation of motion becomes:

$$I\frac{d^2\theta}{dt^2}+kr\frac{d\theta}{dt}+\frac{G}{l}\theta=A\cos pt, \tag{6}$$

where r is the distance from the center of rotation and the solution of this equation (6) is:

$$\theta=e^{\frac{-kr}{2I}t}[C\cos p_0 t+D\sin p_0 t]+\frac{A\cos(pt-\psi)}{I\rho} \tag{7}$$

$$\text{where } \psi=\tan^{-1}\frac{\frac{kp}{I}}{p_0^2-p^2} \text{ and } \rho=\sqrt{(p_0^2-p^2)^2+\frac{k^2r^2p^2}{I^2}}.$$

The first term of the second member of equation (7) represents the transient or free components of the oscillation. These components decrease exponentially with increasing time.

The second term of the second member of equation (7) represents the oscillations in the steady state. These stationary oscillations θ_1, are

$$\theta_1=\frac{A\cos(pt-\psi)}{I\rho}=\frac{(A/I)\cos(pt-\psi)}{\sqrt{(p_0^2-p^2)^2+\frac{k^2r^2p^2}{I^2}}}. \tag{8}$$

This equation (8) readily shows that in the steady state the deformations remain oscillatory with a circular frequency p equal to the circular frequency of the periodic force, and with constant amplitude

$$h=\frac{A/I}{\sqrt{(p_0^2-p^2)^2+\frac{k^2r^2p^2}{I^2}}}.$$

It also results from equation (8) that, with respect to the undamped forced oscillations, the damped forced oscillations have undergone a relative retardation in phase ψ.

If m harmonic terms enter into the disturbing force $f(t)$, equation (8) becomes:

$$\theta_1 = \sum_{\nu}^{1,2,\ldots m} \frac{A_\nu / I \cos(p_\nu t - \psi_\nu)}{\sqrt{(p_0^2 - p_\nu^2)^2 + \frac{k^2 r^2 p_\nu^2}{I^2}}} \cdot \tag{8a}$$

(3) Let the tangential forces F_1, F_2, . . . F_n act successively at their respective points of application. Let them act at regular intervals of time T, although any periodicity for each of them could be readily considered. This may represent a revolving disc with a stationary force, which in turn may be constant or periodic. If the disc and force are both rotating with different angular velocities, one of them may be treated as stationary by considering the relative angular velocity of the disc and the force.

In the case of actual rotation of the disc at constant angular velocity, the stresses due to centrifugal action will be superposed on the dynamic stresses as a static field of force, *i.e.*, independent of the time, and only result in an additional permanent static deformation. In the case of varying angular velocity, the centrifugal forces themselves constitute a non-static field. The problem loses its simplicity and will not be considered at present.

Each of the F_n forces, acting with a circular frequency $\frac{1}{nT}$, will set up oscillatory deformations as expressed by equations (2), (4), (5), (7), (8), and (8a).

By virtue of the principle of superposition, the resultant deformation at any time will be the sum of the individual deformations due to each of the forces acting alone. If any of the deformation components becomes infinite, the resultant deformation becomes infinite.

The cyclical periodic applications of the n forces will result, in the steady state, in an oscillation of the type embodied in equation (8a), except that the summation will extend to all the forces F_1, F_2, . . . F_n acting with a frequency $\frac{1}{nT}$. It can be expressed as follows:

$$\theta_1 = \sum_{\lambda}^{1,2,\ldots n} \sum_{\nu}^{1,2,\ldots m} \frac{(A_{\nu\lambda}/I)\cos(p_{\nu\lambda}t - \psi_{\nu\lambda})}{\sqrt{(p_0^2 - p_{\nu\lambda}^2)^2 + \dfrac{k^2r^2p_{\nu\lambda}^2}{I^2}}}.$$

A few characteristics of these oscillatory stresses and strains are:

1. With no dissipative forces, if $p_0 = p_{\nu\lambda}$, the amplitude becomes infinite. This case of resonance between the natural frequency p_0 of free oscillations of the system and the frequency of one of the disturbing forces, leads, in a frictionless system (also in a massless system where $I = \Sigma mr^2 = 0$), to an infinite stress and infinite strain. The diagram with stress and strain as ordinate against time as abscissa at any given point, excepting points on the axis of rotation, is therefore qualitatively represented by Fig. 2.

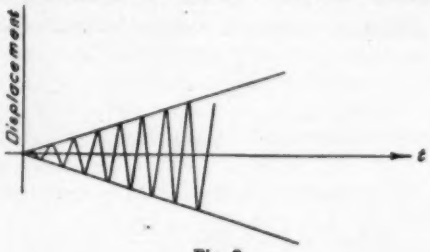


Fig. 2

2. With damping frictional forces, the amplitude still depends upon the frequencies p_0 and $p_{\nu\lambda}$. However, if $p_{\nu\lambda}$ is equal to p_0 the amplitude does not become infinite. The maximum amplitude will correspond to the value of $p_{\nu\lambda}$ for which

$$\rho = \sqrt{(p_0^2 - p_{\nu\lambda}^2)^2 + \frac{k^2r^2p_{\nu\lambda}^2}{I^2}}$$

is a minimum.

This value is readily obtained by equating to zero the derivative of ρ with respect to $p_{\nu\lambda}$:

$$\tfrac{1}{2}\frac{d\rho}{dp_{\nu\lambda}} = \frac{-2(p_0^2 - p_{\nu\lambda}^2)p_{\nu\lambda} + \dfrac{k^2r^2p_{\nu\lambda}}{I^2}}{\sqrt{(p_0^2 - p_{\nu\lambda}^2)^2 + \dfrac{k^2r^2p_{\nu\lambda}^2}{I^2}}} = 0$$

whence $p_{\nu\lambda}^2 = p_0^2 - \frac{k^2r^2}{2I^2}$,

$$(\theta_1)_{\max} = \frac{A_{\nu\lambda}\cos\left[\left(p_0^2 - \frac{k^2r^2}{2I^2}\right)t - \psi_{\nu\lambda}\right]}{kr\sqrt{p_0^2 - \frac{k^2r^2}{4I^2}}}$$

and

$$(h_1)_{\max} = \frac{A_{\nu\lambda}}{kr\sqrt{p_0^2 - \frac{k^2r^2}{4I^2}}}.$$

Dynamic Stresses in pseudo-continuous media.

A. **Free undamped motion in pseudo-continuous media.** We have so far assumed that each particle of the system is free to

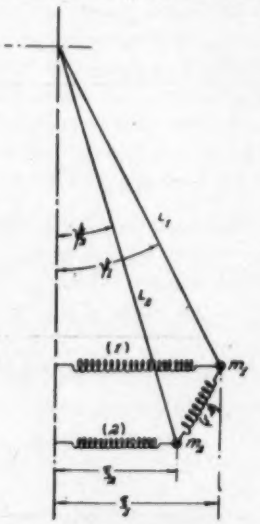


Fig. 3

vibrate irrespective of the state of motion of the neighboring points. Let us now consider the particles of the disc elastically

connected: we express this property by calling the medium a pseudo-continuous medium.

This case can be treated by the Lagrangian equations of motion. Let us consider, Fig. 3, any two points m_1 and m_2 upon which are acting the usual elastic forces (1) and (2) and also the mutual elastic connection (12). Let their position of equilibrium in the unstressed state be $\psi=0$, and Fig. 3 represent their positions at a given time t.

The partial potential energies of this system are:

$$\phi_{12}=\frac{E_1}{2}(\xi_2-\xi_1)^2$$

$$\phi_1=\frac{E_2}{2}\xi_1^2$$

$$\phi_2=\frac{E_2}{2}\xi_2^2$$

where ξ_1 and ξ_2 are the linear displacements and E_1 and E_2 elastic constants.

The total potential energy of the system is:

$$\phi=\frac{E_1}{2}(\xi_2-\xi_1)^2+\frac{E_2}{2}(\xi_2^2+\xi_1^2).$$

The total kinetic energy T is similarly obtained by summation of the partial kinetic energies and is equal to:

$$T=\frac{m_1}{2}\left(\frac{d\xi_1}{dt}\right)^2+\frac{m_2}{2}\left(\frac{d\xi_2}{dt}\right)^2.$$

The equations of motion in the Lagrangian form here being:

$$\frac{d}{dt}\left[\frac{\partial T}{\partial\left(\frac{d\xi_\nu}{dt}\right)}\right]+\frac{\partial\phi}{\partial\xi_\nu}=Q_\nu \tag{9}$$

where T and ϕ are the kinetic and potential energies, ξ_ν the ν independent coördinates necessary to define fully the position of the system, and Q_ν the corresponding generalized disturbing force components.

Making the following substitution of variables:

$$\xi_1 = l_1\psi_1$$
$$\xi_2 = l_2\psi_2$$

We obtain:

$$\phi = \frac{\psi_1^2}{2}(E_1l_1^2+E_2l_1^2)+\frac{\psi_2^2}{2}(E_1l_2^2+E_2l_2^2)-E_1l_1l_2\psi_1\psi_2$$

$$\frac{\partial\phi}{\partial\psi_1} = (E_1l_1^2+E_2l_1^2)\psi_1 - E_1l_1l_2\psi_2$$

$$\frac{\partial\phi}{\partial\psi_2} = (E_1l_2^2+E_2l_2^2)\psi_2 - E_1l_1l_2\psi_1.$$

Also:

$$T = \frac{m_1}{2}l_1^2\left(\frac{d\psi_1}{dt}\right)^2+\frac{m_2}{2}l_2^2\left(\frac{d\psi_2}{dt}\right)^2.$$

If $\frac{d\psi_1}{dt} = X_1$ and $\frac{d\psi_2}{dt} = X_2$

then: $$T = \frac{m_1}{2}l_1^2X_1^2+\frac{m_2}{2}l_2^2X_2^2$$

and: $$\frac{d}{dt}\left(\frac{\partial T}{\partial X_1}\right) = m_1l_1^2\frac{d^2\psi_1}{dt^2}\ ,\quad \frac{d}{dt}\left(\frac{\partial T}{\partial X_2}\right) = m_2l_2^2\frac{d^2\psi_2}{dt^2}.$$

Introducing these values in Equation (9), Q_r being equal to zero, we obtain the following differential equations of motion:

$$\left.\begin{aligned} m_1l_1^2\frac{d^2\psi_1}{dt^2}+(E_1l_1^2+E_2l_1^2)\psi_1-E_1l_1l_2\psi_2 = 0 \\ m_2l_2^2\frac{d^2\psi_2}{dt^2}+(E_1l_2^2+E_2l_2^2)\psi_2-E_1l_1l_2\psi_1 = 0 \end{aligned}\right\} \qquad (10)$$

The solution of these equations leads to the kinetic equations of motion $\psi_1 = f_1(t)$ and $\psi_2 = f_2(t)$ of the two particles m_1 and m_2 allowed to vibrate under the sole action of elastic connections.

A solution of these two linear differential equations of the second order is obtained by writing:

$$\left.\begin{aligned} \psi_1 = A_1e^{ipt} \\ \psi_2 = A_2e^{ipt} \end{aligned}\right\} \qquad (11)$$

where A_1, A_2 and p are three parameters which must be determined.

Differentiating (11) twice with respect to t and substituting in (10) ψ_1, ψ_2, $\frac{d^2\psi_1}{dt^2}$, $\frac{d^2\psi_2}{dt^2}$, by the values obtained, equations (10) become:

$$\left.\begin{aligned} A_1[E_1l_1^2+E_2l_1^2-m_1l_1^2p^2]+A_2[-E_1l_1l_2]=0 \\ A_1[-E_1l_1l_2]+A_2[E_1l_2^2+E_2l_2^2-m_2l_2^2p^2]=0 \end{aligned}\right\} \tag{12}$$

Neglecting the trivial solution $A_1=A_2=0$, these two linear equations will be generally satisfied under the condition that the determinant formed with the coefficients of A_1 and A_2 be equal to zero. Whence:

$$\begin{vmatrix} (E_1l_1^2+E_2l_1^2-m_1l_1^2p^2), & -E_1l_1l_2 \\ -E_1l_1l_2, & (E_1l_2^2+E_2l_2^2-m_2l_2^2p^2) \end{vmatrix}=0$$

and

$$Ap^4-Bp^2+C=0 \tag{13}$$

where

$$\begin{aligned} A&=m_1m_2l_1^2l_2^2 \\ B&=l_1^2l_2^2[(m_1+m_2)\ (E_1+E_2)] \\ C&=l_1^2l_2^2(E_2^2+2E_1E_2) \end{aligned}$$

The roots of equation (13) being:

$$\left.\begin{aligned} p_1^2&=\frac{B+\sqrt{B^2-4AC}}{2A} \\ p_2^2&=\frac{B-\sqrt{B^2-4AC}}{2A} \end{aligned}\right\} \tag{14}$$

where p_1 and p_2 are the circular frequencies of m_1 and m_2. If the symmetric case, in which

$$\begin{aligned} m_1&=m_2=m \\ l_1&=l_2=l \\ E_1&=E_2=E \end{aligned}$$

is considered, the frequencies become:

$$p_1^2 = \frac{3E}{m}$$

$$p_2^2 = \frac{E}{m}$$

and

$$\psi_1 = A_1 \cos \sqrt{\frac{3E}{m}}\, t \qquad (15)$$

$$\psi_2 = A_2 \cos \sqrt{\frac{E}{m}}\, t.$$

For any positive value of m_1 and m_2, p_1 and p_2 will not be equal. It is thereby seen that the different points of the pseudo-continuous medium are not oscillating in synchronism.

B. **Forced damped oscillations in pseudo-continuous media.** Equations (15) are the solutions for a free non-dissipative system. Let us now consider this same system under the influence of a periodic force — which may be expressed as a harmonic function of the time and consequently as the real part of an imaginary exponential function of the time — and under the influence of elastic and dissipative forces respectively proportional to the displacement and to the velocity.

The motion of the system is then given by the two following differential equations for which the Heaviside operational method[3] offers a solution.

$$\left.\begin{aligned} m_1 l_1^2 \frac{d^2\psi_1}{dt^2} + (E_1 l_1^2 + E_2 l_1^2)\psi_1 + k l_1 \frac{d\psi_1}{dt} - E_1 l_1 l_2 \psi_2 &= f(t) \\ m_2 l_2^2 \frac{d^2\psi_2}{dt^2} + (E_1 l_2^2 + E_2 l_2^2)\psi_2 + k l_2 \frac{d\psi_2}{dt} - E_1 l_1 l_2 \psi_1 &= 0 \end{aligned}\right\} \qquad (16)$$

[3] O. Heaviside — Elec. Pap. Vol. II, p. 371.
An elementary presentation is given in "Notes on Heaviside's Operational Method" by M. S. Vallarta, M. I. T., March, 1923.

Operationally, equations (16) may be written in the form:

$$\left.\begin{aligned}(a_{11}p^2+b_{11}p+c_{11})\psi_1+(a_{12}p^2+b_{12}p+c_{12})\psi_2=f(t)\\(a_{21}p^2+b_{21}p+c_{21})\psi_1+(a_{22}p^2+b_{22}p+c_{22})\psi_2=0\end{aligned}\right\} \quad (17)$$

where p^n is the operator $\frac{d^n}{dt^n}$ which, when applied to an operand ψ yields the n^{th} derivative of ψ with respect to the parameter t, n being a real integer.

In general the disturbing force will be expressed by an absolutely convergent series of harmonic terms. If more than one harmonic term is considered, the solution of equations (17) has to be written for each harmonic force component separately and the resultant motion — by virtue of the principle of superposition, or of the principle of independent action of forces — is obtained as before by the addition of the individual displacements due to each force component.

The solutions of the primitive equations (16) will again consist of two parts, the forced components of oscillation and the free components. As before, the forced components are obtained from the particular solutions of the equations and their frequency will be the frequency of the disturbing force.

These particular solutions are obtained in the familiar way by solving the above system of linear equations (17). Denoting them by y_1 and y_2 we have:

$$y_1=\frac{\begin{vmatrix}Ae^{ipt}, & a_{12}p^2+b_{12}p+c_{12}\\ 0, & a_{22}p^2+b_{22}p+c_{22}\end{vmatrix}}{\begin{vmatrix}a_{11}p^2+b_{11}p+c_{11}, & a_{12}p^2+b_{12}p+c_{12}\\ a_{21}p^2+b_{21}p+c_{21}, & a_{22}p^2+b_{22}p+c_{22}\end{vmatrix}}=\frac{\Delta_{22}(p)}{\Delta(p)}Ae^{ipt} \quad (18)$$

and likewise

$$y_2=\frac{\Delta_{21}(p)}{\Delta(p)}Ae^{ipt}$$

where

$$a_{11}=m_1l_1^2 \qquad a_{22}=m_2l_2^2$$
$$b_{11}=kl_1 \qquad b_{22}=kl_2$$
$$c_{11}=E_1l_1^2+E_2l_1^2 \qquad c_{22}=E_1l_2^2+E_2l_2^2$$
$$c_{12}=-E_1l_1l_2 \qquad c_{21}=-E_1l_1l_2$$
$$a_{12}=b_{12}=a_{21}=b_{21}=0$$

The free components z_1 and z_2, given by the solutions of the *reduced* equations (17), are:

$$\left.\begin{aligned} z_1&=\sum_k B_{1k}e^{ip_kt} \\ z_2&=\sum_k B_{2k}e^{ip_kt} \end{aligned}\right\} \tag{19}$$

where p_k is a root of the determinantal equation $\Delta(p)=0$ and the summation for z_1 and z_2 is extended over the degrees of freedom of each particle

Thereby the complete solutions of equations (16) are:

$$\left.\begin{aligned} \psi_1&=y_1+z_1=\frac{\Delta_{22}(p)}{\Delta(p)}Ae^{ipt}+\sum_k B_{1k}e^{ip_kt} \\ \psi_2&=y_2+z_2=\frac{\Delta_{21}(p}{\Delta(p)}Ae^{ipt}+\sum_k B_{2k}e^{ip_kt} \end{aligned}\right\} \tag{20}$$

Carrying solutions (20) further for the symmetrical conditions where

$$a_{11}=a_{22}=ml^2$$
$$b_{11}=b_{22}=kl$$
$$c_{11}=c_{22}=2El^2$$
$$c_{12}=c_{21}=-El^2$$
$$a_{12}=b_{12}=a_{21}=b_{21}=0$$

we obtain:

$$\Delta_{22}(p)=\begin{vmatrix} 1, & -El^2 \\ 0, & ml^2p^2+klp+2El^2 \end{vmatrix}=(ml^2p^2+klp+2El^2) \tag{21}$$

$$\Delta_{21}(p)=2El^2 \tag{22}$$

$$\begin{aligned}\Delta(p) &=(ml^2p^2+klp+2El^2)^2-E^2l^4 \\ &=m^2l^4p^4+2kl^3mp^3+(4Eml^4+k^2l^2)p^2+4Ekl^3p+3E^2l^4. \end{aligned} \tag{23}$$

Let α_1, α_2 be the roots of equations (21), α_3, α_4, α_5, α_6 the roots

of equation (23) when equated to zero. Equations (20) become:

$$\left.\begin{aligned}\psi_1 &= \frac{(p-a_1)(p-a_2)}{(p-a_3)(p-a_4)(p-a_5)(p-a_6)}\, Ae^{ipt}+B_{11}e^{ia_4t}+B_{12}e^{ia_6t}\\ \psi_2 &= \frac{2El^2}{(p-a_3)(p-a_4)(p-a_5)(p-a_6)}\, Ae^{ipt}+Be^{ia_4t}+Be^{ia_6t}\end{aligned}\right\} \quad (24)$$

or:

$$\left.\begin{aligned}\psi_1 &= \frac{(p-a_1)(p-a_2)}{(p-a_3)(p-a_4)(p-a_5)(p-a_6)}\, A\cos pt+B_{11}\cos a_4t+B_{12}\cos a_6t\\ \psi_2 &= \frac{2El^2}{(p-a_3)(p-a_4)(p-a_5)(p-a_6)}\, A\cos pt+B_{21}\cos a_4t+B_{22}\cos a_6t\end{aligned}\right\} \quad (25)$$

where p is the circular frequency of the disturbing force. The first term of the second members represents the forced components of the steady state and the following terms represent the transient components. Similarly as in equation (14), equations (25) show that the transient oscillations of the different parts of the system have different frequencies: the resulting relative variation in phase and amplitude from point to point merely expresses the existence of elastic waves.

If one of the roots of the determinantal equation (23) approaches the circular frequency of the disturbing force, the amplitude of the steady state oscillations approaches infinity.

For simplicity the argument has been restricted to a system having two degrees of freedom, but the method can be applied to systems of this type with a finite number of degrees of freedom.[4] The general solutions are still represented by equations of the form of equation (20), the determinantal equation $\Delta(p)=0$ simply increasing in order.

It must be noted that the stress and strain components, entering into the determination of the dynamic states of stress and the dynamic states of strain throughout the medium cannot in general be arbitrarily chosen. They must fully define the elastic state at each point and be propagated with the same velocity. Among the different stress and strain components only the pure shear

[4] J. R. Carson, Phys. Rev. X, 217, 1917.

or rotational components and the pure dilatational components are known to fulfill the second condition. The pure rotational components, however, do not suffice to define fully the elastic state whereas the pure dilatational components under the form of the three principal stresses or the three principal dilatations, when known in direction and in magnitude, completely determine at each point the state of stress and the state of strain. They travel through a continuous isotropic elastic medium and consequently also through a pseudo-continuous isotropic elastic medium, in a direction normal to their respective isostatic surfaces, with a velocity equal to $\sqrt{\frac{\lambda+2\mu}{\rho}}$ where ρ is the density, λ and μ Lamé's classical elastic constants.

The above theory, neglecting the reflection of the elastic waves, tacitly postulates an infinitely extending medium. In the majority of physical cases the dissipation will be great enough so that the amplitude of the waves reflected at the boundaries of the finite medium and the interference effects, which would result, may be neglected. This theory can then without further development be applied to media whose boundaries lie at finite distances.

Contribution No. 5
Research Laboratory of Theoretical Physics
April, 1924